

Do Personality and Culture Influence Perceived Video Quality and Enjoyment?

Michael James Scott*, Sharath Chandra Guntuku*,
Weisi Lin *Senior Member, IEEE*, Gheorghita Ghinea *Member, IEEE*,

Abstract—The interplay between system, context and human factors is important in perception of multimedia quality. However, studies on human factors are very limited in comparison to those for system and context factors. This article presents an attempt to explore the influence of personality and cultural traits on perception of multimedia quality. As a first step, a database consisting of 144 video sequences from 12 short movie excerpts has been assembled and rated by 114 participants from a cross-cultural population. Thereby providing a useful ground-truth for this (as well as future) study. As a second step, three statistical models are compared: (i) a baseline model to only consider system factors; (ii) an extended model to include personality and culture; and (iii) an optimistic model in which each participant is modeled. As a third step, predictive models based on content, affect, system, and human factors are trained to generalize the statistical findings. As shown by statistical analysis, personality and cultural traits represent 9.3% of the variance attributable to human factors and human factors overall predict an equal or higher proportion of variance compared to system factors. Moreover, the quality-enjoyment correlation varies across the excerpts. Predictive models trained by including human factors demonstrate about 3% and 9% improvement over models trained solely based on system factors for predicting perceived quality and enjoyment. As evidenced by this, human factors indeed are important in perceptual multimedia quality, but the results suggest further investigation of moderation effects and a broader range of human factors is necessary.

Index Terms—Multimedia; Quality; QoE; Personality; Big-5; Culture; Hofstede; Enjoyment; Perception; Video

I. INTRODUCTION

Multimedia quality is a multi-faceted concept, dependent on a) the nature of content (i.e., affective narrative), b) the parameters controlling how that content is delivered, and c) the individual traits of users which influence their perception of quality. The nature of content can only be manipulated by content creators, however providers have the ability to stream the content at the ‘right’ parameters to ensure adequate viewer satisfaction. A user perspective of multimedia quality is both necessary and challenging. It is necessary for the simple reason that users have a multitude of multimedia-based contents and services to choose from. If their perceived

quality is poor, users have a variety of competitor offerings to go for instead. Focusing on purely system factors alone, ignoring the user, also has the danger that redundancies beyond the ones currently incorporated in audio and video compression schemes will go unexploited. For instance, if one takes into account user task or context of use, perceptions of what constitutes good multimedia quality may be lower than what would usually be expected. In other words, current audio-video compression algorithms tend to over-estimate resources required to achieve good multimedia quality. This, in turn, may lead to inefficient resource utilization.

Suppose that two viewers are both shown a video sequence each at {25fps, 480p} and {15fps, 720p}. Is it possible that one viewer’s perceived quality on a video sequence with a lower parameter setting is broadly equivalent to the other viewer’s perceptive quality on another sequence with higher parameters? If yes, what are the factors influencing perception of quality? Which of these factors contribute positively and which negatively? These are some of the questions which motivate the research discussed in this paper. Finding answers to these questions would help content providers to maintain adequate customer satisfaction in a personalized manner while optimally utilizing resources such as bandwidth.

The challenge of adopting a user-perspective of multimedia quality is that, in incorporating subjective tests, it is a costly exercise in terms of effort and resources (not least human) needed. In the long run, however, the benefits of adopting a user perspective of multimedia will more than compensate for its associated cost, and there have been a growing awareness and research efforts devoted to perceptual multimedia quality over the past decade. Most of these fall under the umbrella of Quality of Experience (QoE) studies and explore user-centric issues pertaining to multimedia audio-visual quality assessment [23], [36], [67], the influence of varying network conditions and transmission models on QoE [54], [31], [64], mapping between QoE and QoS (Quality of Service) [10], on emotional, sensorial, and cognitive responses associated with multimedia content [14], [16], [52], as well, as more recently, of using crowd-sourcing for QoE [27], [65]. However, with a few exceptions [13], [70], such research has largely neglected individual user traits. This is a critical omission because of the potential advantage of exploiting such traits through, for instance, providing personalized content and delivery.

This is precisely the niche that is explored in this paper. Accordingly, recognizing that individual traits can be explored both at macro-cosmic (e.g. cultural influences) and

Michael James Scott is with the Games Academy, Falmouth University, Cornwall, United Kingdom E-mail: michael.scott@falmouth.ac.uk.

Sharath Chandra Guntuku and Weisi Lin are with School of Computer Engineering, Nanyang Technological University, Singapore, 639798 E-mail: sharathc001@e.ntu.edu.sg, wslin@ntu.edu.sg.

Gheorghita Ghinea is with Department of Computer Science, Brunel University, London, United Kingdom E-mail: george.ghinea@brunel.ac.uk.

Manuscript received .

* denotes equal contribution

micro-cosmic (e.g. personality type) levels, this article presents the results of a twin-center study run in Singapore and the United Kingdom which investigates the relationship between user perceived multimedia quality, personality, and cultural traits. Based on the results, several predictive models for user perceived quality and enjoyment are derived.

The structure of this paper is as follows: Section II presents related work, while Section III presents the data collection process, the methodology used to generate the predictive models is then detailed in Section IV, whilst the results of their application to our dataset is the subject of Section V. The implications of our work are discussed in Section VI, with conclusions being drawn and opportunities for future work identified in Section VII.

II. RELATED WORK

A. On Video Quality

The last few years have seen many studies on video quality and enjoyment evaluation. To enable the same, numerous video datasets have been created. To name a few - VQEG HDTV database [19]; LIVE video database [58]; IVC video databases [37]; ReTRiEVED video database [50]; video enjoyment database [38], and aesthetic evaluation video database [46]. Human subjects evaluated the visual quality of videos in these datasets, by inspecting the videos derived from different distortion types (e.g., H.264 compression, packet loss and frame rate change). There were also works in the quality assessment domain, starting with building basic multimedia quality model [23] and others focusing on foveated and compressed images and video streams [36], for low bit-rate videos [64], [67], multimedia transmitted over packet networks [54], over UMTS networks [31], for H.264/AVC coded videos [48], on 3D videos [32], for mobile videos [60] etc. There have also been works on crowdsourcing QoE [27], [65], subjective quality evaluation via paired comparison [34], [35]. However, the role played by human factors was not thoroughly investigated, leaving a lot of scope to study their influence on multimedia quality. A similar approach is followed in aesthetic evaluation datasets to obtain the ground-truth for enjoyment (or appeal) on different videos. However, only four human factors (namely age, gender, vision and expertise levels) are reported. This leaves open an assumption that all participants have the same or similar perception to the visual quality of a video, irrespective of other human factors.

B. On Personality and Culture

The systematic differences in individuals' traits can be explored using personality, which is defined as a series of "internal properties" that relate to overt behaviors [42]. The Five Factor Model (FFM) [15] is one of the most used theories to examine the predictive utility of personality. This consists of: openness to experience, conscientiousness, extroversion, agreeableness and neuroticism. These dimensions could map to perception of quality in many ways. For example, individuals with high neuroticism may be more sensitive to multimedia content that evokes negative emotions.

Apart from traits which vary across a wide range of individuals, there are also traits which are associated with the culture which the individual is from. The local environment in which one lives and grows also shape one's perception and cognition through shared conceptions and collective norms [26]. Hofstede's Six Factor Model (HM) [24] studies these cultural differences. The model includes six cultural traits: power distance; individualism; uncertainty avoidance index; masculinity; pragmatism; and indulgence. Again, each of these dimensions could interact with the perception of quality and enjoyment. For example, individuals with high indulgence may become more critical of due to extended usage.

C. Background

Previous educational and socio-cultural backgrounds are shown to play important roles in subjective ratings [16], [55]. The influence of cultural experience has been shown in visual perception while viewing objects. For instance, in [49], it was found that there exists perceptual and attentional differences between Asians and Westerners. In that study, Americans were reported to have more analytical visual perception (inclined to pay attention to details), while Asians were seen to have a more holistic visual perception (likely to be more sensitive to context). In [44], a more detailed investigation on the cultural differences in cognitive processes can be found. The influence of culture on optical illusion, color perception, visual attention and brain functioning was studied in [17]. The correlation between culture and cognition was also studied in [1], by analyzing the variation of word associations given by Japanese and American participants.

Factors which influence perception of quality and emotion were also studied previously (e.g. [10], [14], [52] and [11], [40] respectively). In [8], [21], the influence of personality on media content preferences has been studied.

As such, previous experience, socio-cultural background, goals, and values influence one's understanding, interpretation and evaluation of. This may, consequently, influence what viewers consider to be of 'good' or 'bad' quality. Some works (for example [12], [28], [41], [52], [53], [59]) investigate the influence of above-mentioned human factors. But they are often based on subjective tests applied to samples. However, it is seldom the case that samples are deliberately drawn from cross-cultural contexts.

Recently, there have been works studies on the influence of personality factors and social context on perceived quality [63], [69]. However, [69] involved 59 users and their ratings on 6 YouTube videos in three genres and [63] focused on investigating the use of a multi-modal remote control application in the context of IPTV. As stated previously, such studies tend to draw their samples just from only the local population.

Hence, it is meaningful to investigate the influences of individual and cultural variation with respect to the perceived quality and subsequent enjoyment video sequences, with a larger group of users from multiple demographic regions using a larger dataset, with various quality parameters. It is important to involve participants from many different countries

TABLE I: Estimated Marginal Means of the Ratings for Each Movie Excerpt in the CP-QAE-I (z-Scores)

Movie Excerpt	Perceived Quality	Enjoyment
A_FISH_CALLED_WANDA	-.300	-.004
AMERICAN_HISTORY_X	-.121	-.560
CHILDS_PLAY_II	-.430	-.181
COPYCAT	-.022	-.443
DEAD_POETS_SOCIETY_1	-.105	-.437
DEAD_POETS_SOCIETY_2	.365	.782
FOREST_GUMP	.448	.747
SE7EN_1	.073	.402
SE7EN_3	-.229	-.312
SOMETHING_ABOUT_MARY	.220	.484
THE_PROFESSIONAL	.131	.330
TRAINSPOTTING	-.013	-.621

Covariates: Ext = 5.54; Agr = 7.22; Con = 6.55; Neu = 5.62; Ope = 6.75; PDI = -35.96; IDV = 18.73; MAS = -1.23; UAI = 44.61; PRG = 16.84; IVR = -16.97.

(as detailed in Section III-B) in the subjective testing where culture and personality could potentially explain differences.

III. DATA COLLECTION

A. Sample Size

Minimum sample size was determined by power analysis, using G*Power 3. Using the conventional error probabilities ($\alpha = .05$, $1 - \beta = 0.8$) and assuming ‘medium’ effect sizes will be detected ($f = 0.39$) and that repeated measures will be correlated ($r = 0.8$), a minimum sample size of 64 was suggested. Due to the risk of error inflation associated with testing a large number of parameters, this was increased to 112. A sample size of 114 was obtained.

B. Participants

The participants were 114 university students drawn from [removed for peer review]. Exactly 50% of the sample was drawn from each institution. In terms of nationality, there were: 43 British, 22 Indian, 16 Chinese, 15 Singaporean, 4 Nigerian, 2 Indonesian, 2 Pakistani, 2 Vietnamese, 1 Danish, 1 Dutch, 1 Latvian, 1 Myanmarian, 1 Polish, 1 Tanzanian, 1 Turkish, and 1 Zimbabwean. The proportion of female participants was 28.9% and the average age was 23.9 years ($\sigma = 3.68$).

Further descriptive statistics including personality and cultural traits across the two institutions involved are shown above in Table II. Responses from participants from both universities are segregated and the mean values of different traits in the individual cohorts are reported in columns 1 and 2. In columns 3 and 4, the mean values and standard deviations of different traits of the combined cohort (pool) are reported. As would be expected, the personality variables follow the normal distribution and are consistent across the institutions. However, differences can be observed with respect to the cultural variables. These include: individualism; masculinity; uncertainty avoidance; and indulgence. It is important to note, however, that the full range of possible values were

TABLE II: Sample Descriptives

Human Factors	$\bar{x}(NTU)$	$\bar{x}(BUL)$	$\bar{x}(Pool)$	σ
Extroversion	5.61	5.46	5.54	1.689
Agreeableness	7.33	7.31	7.22	1.533
Conscientiousness	6.40	6.70	6.55	1.523
Neuroticism	5.56	5.68	5.62	1.716
Openness	6.60	6.91	6.75	1.424
Power Distance	-35.61	-36.32	-35.96	53.219
Individualism	25.79	11.67	18.73	50.619
Masculinity	3.68	-6.14	-1.23	53.483
Uncertainty Avoidance	52.54	36.67	44.61	47.182
Pragmatism	16.14	17.54	16.84	58.090
Indulgence	-22.63	-11.32	-16.97	65.522

\bar{x} : Sample Mean; σ : Standard Deviation

not observed for: extroversion (87.5% of expected range); openness (75% of expected range); and masculinity (67% of expected range).

A non-probability sampling method was used to recruit the participants. This tends to have two key weaknesses: lack of prototypicality (i.e., does the sample represent the target population?); and range restriction (i.e., is there sufficient variance in variables of interest to detect a relationship?). However, the focus of this study is modeling, rather than demography. Thus, only the latter concern presents a potential threat to validity. With the exception of those variables for which a full range of responses were not observed, there is little evidence of range restriction for most of the key variables of interest in this study.

C. Video Dataset

The CP-QAE-I video dataset (available from <http://1drv.ms/1M1bnwU>) was validated in a previous study [22], and the study presented in this article represents its first full-scale application. The video dataset contains sequences based on 12 excerpts from popular movies that were selected purposively to evoke different affects [57]. Movie clips of different valence, and which had least variation in ratings on arousal [23], were considered for this study to minimize any content-based biases (a list of the clips along with mean affect scores are seen in Table I). Three parameters were also varied: bit-rate (384kb/s and 768kb/s); frame dimension (480p and 720p); and frame rate (5fps, 15fps and 25fps).

Thus, in the CP-QAE-I, there were 144 videos (resulting from the 3*2*2*12 conditions of varying frame-rate, frame dimension, bit-rate, and excerpt). Each video sequence has a length between 1 and 3 minutes. The estimated marginal means demonstrate sufficient distinction between the video sequences in terms of perceived quality and enjoyment as shown above in Table I.

D. Measures

Several measurement scales were used in the study to capture data from participants with respect to: perception

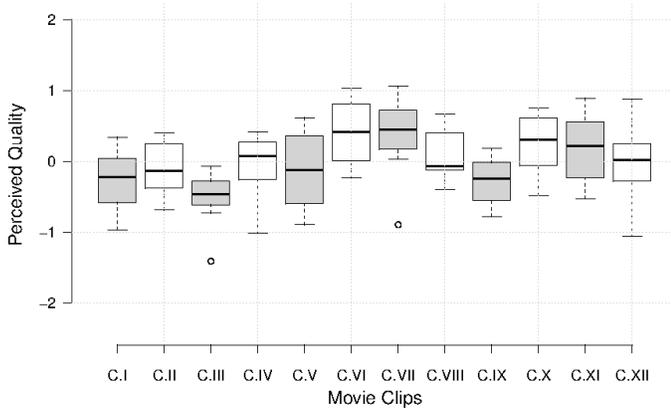


Fig. 1: Box-Plot showing the distribution of Perceived Quality of Each Movie Excerpt in the Dataset (z-Score)

of quality; enjoyment; culture; and personality. All were previously validated and defined for participants.

1) *Quality*: The QoP-LoQ scale [20] was used to measure subjective perception of quality. This is a single 5-point Likert-type rating item where participants indicate satisfaction with the quality of the video sequence. A high score indicates “absolute” satisfaction while a low score indicates “no” satisfaction.

2) *Enjoyment*: A subjective measure was deployed to measure enjoyment. This was a single 5-point Guttman-type rating item where participants indicate their level of enjoyment. A high score indicates “high” enjoyment, but a low score indicates “no” enjoyment.

3) *Culture*: Culture is collective, rather than individual, so when measured at the individual-level the traits associated with culture are what is being assessed. The VSM-2013 questionnaire [25] was used to measure these traits, following Hofstede’s dimensions: power distance (PDI); individualism (IDV); uncertainty avoidance (UAI); masculinity (MAS); pragmatism (PRG); and indulgence (IVR).

4) *Personality*: The BFI-10 [18] questionnaire was used to measure personality. The questions required a Likert-type response based on the FFM [15]: openness (Ope); conscientiousness (Con); Extroversion (Ext); Agreeableness (Agr); Neuroticism (Neu).

E. Procedure

A lab-based testing approach was adopted. Participants began by responding to the VSM-2013 [25] and the BFI-10 [18] questionnaires to report cultural and personality traits, respectively. Then each participant watched 14 video sequences: 2 training video sequences (one at the beginning and a reminder at the halfway point to show participants what should be considered ‘high’ quality); and 12 videos under assessment. Participants were randomly allocated quality parameters for each individual video sequence (except for the training videos). Because we were studying the effects of personality and culture, we wanted to maximise the ecological validity of participants viewing behaviour. Therefore, we did not control the audio volume, ambient light/noise. Screen

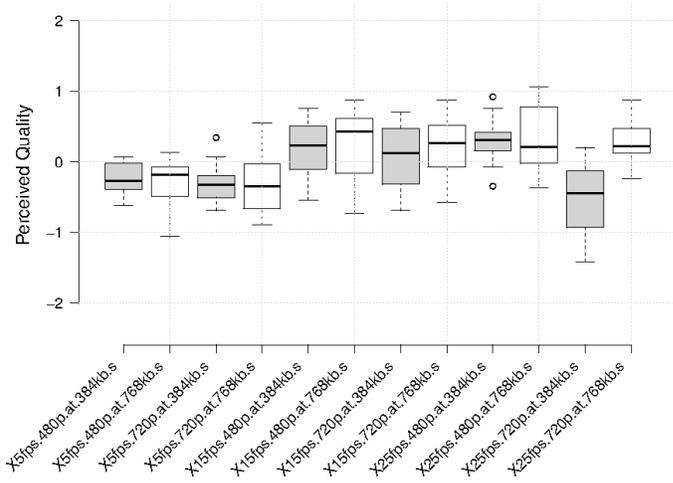


Fig. 2: Box-Plot showing the distribution of Perceived Quality of Each Parameter Setting in the Dataset (z-Score)

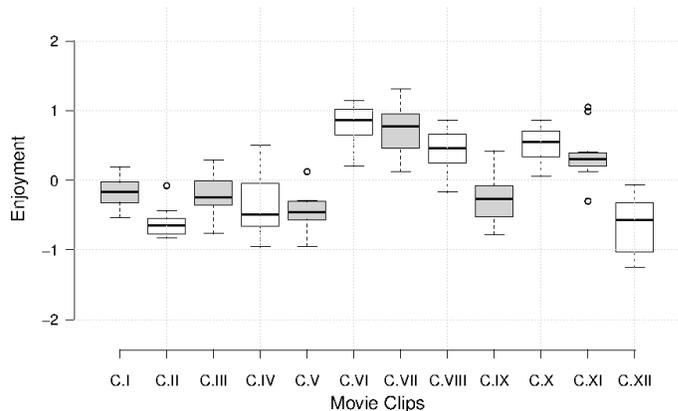


Fig. 3: Box-Plot showing the distribution of Enjoyment of Each Movie Excerpt in the Dataset (z-Score)

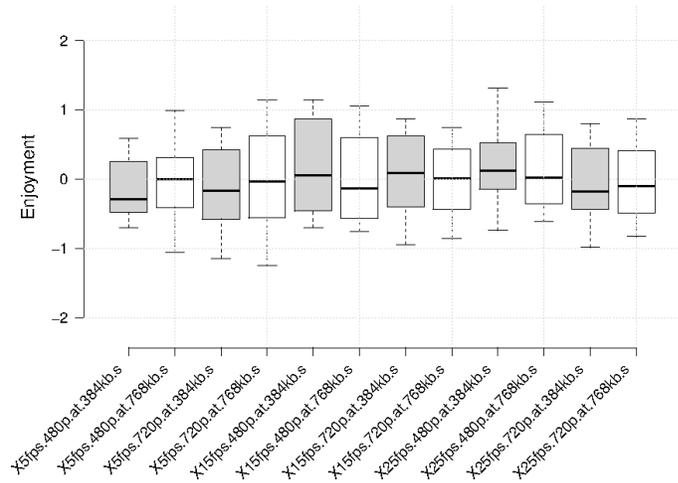


Fig. 4: Box-Plot showing the distribution of Enjoyment of Each Parameter Setting in the Dataset (z-Score)

size (23") and audio quality were consistent. The participants were left free and without any time-limit to complete the experiment. Furthermore, the video sequences were hosted on a web server locally at each institution involved in the study to minimize latency.

Ethical considerations, including anonymity and informed consent, were assured throughout the study. Subsequently, while each participant was expected to rate all 12 video sequences, only 73.7% of participants actually rated all 12. The minimum number of videos rated was 3, however the average was 10.8 ($\sigma = 2.56$). In total, 1232 ratings were recorded, which corresponds to 90% of the maximum possible.

IV. DATA ANALYSIS METHODS

A. Statistical Analysis

PASW 18.0.3 for Windows was used to conduct the analyses. Due to the diversity of nationalities involved in the study, as well as the overlap between the nationality of participants each institution recruited, the geographic location of the institutions could not be used as a proxy for culture. Hence, they are not compared. Instead, cultural variables were captured at the individual-level and regression models are used to examine key differences in terms of these variables. Mixed linear regression has been used to account for repeated measures. Parameters were estimated concurrently using the restricted maximum-likelihood method. Missing data was excluded pair-wise.

1) *Baseline Model*: Only system factors are considered in the baseline model. As such, there were 12 variations of system factors which are considered within the context of the CP-QAE: frame-rate (3 conditions); frame dimension (2 conditions); and bit-rate (2 conditions). Other important factors, such as file format and delivery protocol, were not varied as part of the experimental design. As interactions between the system factors is expected (e.g., minimizing bit-rate while maximizing other system factors would likely cause visual artifacts), these were modeled as factorial interactions. Furthermore, the movie excerpts are considered a parameter due to differences in cinematographic technologies and techniques.

2) *Extended Model*: The baseline model is then extended with several additional fixed parameters which are included as covariates. These are the personal and cultural traits considered in this study, including: extroversion; agreeableness; conscientiousness; neuroticism; openness; power distance; individualism; masculinity; uncertainty avoidance; pragmatism; and indulgence.

3) *Optimistic Model*: The goal of a model is to predict the value of a dependent variable accurately, however it is often the case that residual variance can arise for a variety of reasons beyond the accuracy of the model. Some of these are: random error; measurement error; and model technique limitations (in this case, the regression only considers linear relationships). As such, the residual variance cannot be attributed to human factors which have not been included in the model. So, to provide an estimate of the proportion of residual variance which could be reasonably attributed to human factors (and,

to some extent, context, due to the limits of experimental control), a third 'optimistic' model is presented. In this model, each participant is a "random effect" such that repeated measurements are used to vary the intercept of the regression for each individual participant.

B. Low System Parameters

A key area of interest for the impact of human factors are conditions where multimedia streaming resources are heavily constrained. For this reason, an analysis was conducted focusing solely on instances where videos with low system parameter settings received high scores for enjoyment and perceived quality. The following system parameter settings were defined as low: a frame dimension of 480p, a frame rate of 5fps, and a bit rate of 384kbps. This filtered subset of the data (about 9.5% of the total ratings) was then investigated using t-tests, comparing those videos which received 'low' ratings (i.e., 1, 2, or 3) with those video sequences which received 'high' ratings (i.e., 4 or 5).

C. Further Predictive Modeling

Initial results suggested that perception of quality may also depend on video characteristics, with inconsistent correlations between perceived quality and enjoyment across different movie excerpts. Hence, a prediction framework which further extended the proposed models with input features related to video content were proposed. The prediction framework is based on the L-1 regularized L2-loss sparse support vector classification using linear kernel. This method was selected to overcome the problem of over-fitting. From each video, frames are extracted using ffmpeg. Image features are extracted from each frame and the average feature of all the frames is taken as the representation for each video.

1) *Features*: Four sets of features (namely content, system, affect and human) were incorporated to represent the different factors that contribute to perceptual quality and enjoyment, as follows:

Color Histogram in RGB space is evaluated, as color is an important cue which influences users' perception, representing users' inclination towards different colors.

Aesthetic Features: Two sets of aesthetic features are used: one proposed by [39], which is used to characterise photographic styles (e.g. rule-of-thirds, vanishing points, etc.) based on art theory and psychological studies, and another proposed by [3], which are psycho-visual statistics extracted from multiple levels, namely cell level, frame level and shot level. These are supposed to give an idea about users' aesthetic preferences.

LBP is used to encode visual texture perception information. Many of the videos include images of people and LBP features represent facial information well.

Bag-of-Visual-Words [66]: A vocabulary is generated by vector quantization of key-point descriptors, where every image is represented as a bag of visual words by mapping the key-points to visual words, instead of encoding photographic rules explicitly.

TABLE III: Baseline Model for Perception of Quality and Enjoyment

Parameter	df_{num}	Perceived Quality			Enjoyment		
		df_{den}	F	p	df_{den}	F	p
Movie Excerpt	11	191.387	8.880	.000	177.090	40.140	.000
Frame Rate (FR)	2	1152.788	23.540	.000	1131.230	5.173	.006
Frame Dimension (Dim)	1	1164.451	16.890	.000	1146.390	2.846	.092
Bit-Rate (BR)	1	1160.518	9.830	.002	1139.690	.474	.491
FR * Dim	2	1150.910	3.070	.047	1130.961	1.663	.190
FR * BR	2	1152.330	5.188	.006	1131.496	2.078	.126
Dim * BR	1	1165.993	8.240	.004	1137.742	1.364	.243
FR * Dim * BR	2	1154.080	5.714	.003	1130.448	.002	.998

TABLE IV: Extended Model for Perception of Quality and Enjoyment

Parameter	df_{num}	Perceived Quality			Enjoyment		
		df_{den}	F	p	df_{den}	F	p
Movie Excerpt	11	191.490	9.070	.000	171.956	39.733	.000
Frame Rate (FR)	2	1142.880	24.075	.000	1136.577	4.695	.009
Frame Dimension (Dim)	1	1153.771	13.578	.000	1151.402	3.336	.068
Bit-Rate (BR)	1	1148.206	12.677	.000	1145.171	.257	.612
FR * Dim	2	1145.057	3.748	.024	1145.206	1.057	.348
FR * BR	2	1144.258	5.262	.005	1138.177	1.856	.157
Dim * BR	1	1154.877	9.876	.002	1146.873	2.424	.120
FR * Dim * BR	2	1146.555	5.981	.003	1138.844	.057	.945
Extroversion	1	1151.392	.130	.718	1150.401	.024	.877
Agreeableness	1	1151.909	2.672	.102	1152.475	2.001	.157
Conscientiousness	1	1141.817	7.126	.008	1141.249	5.271	.022
Neuroticism	1	1149.100	11.708	.001	1146.479	.050	.823
Openness	1	1150.056	1.168	.280	1145.365	4.344	.037
Power Distance	1	1154.125	.290	.590	1152.465	9.138	.003
Individualism	1	1149.721	5.519	.019	1150.026	.674	.412
Masculinity	1	1147.422	5.578	.018	1141.312	3.312	.069
Uncertainty Avoidance	1	1144.686	.333	.564	1144.106	5.751	.017
Pragmatism	1	1152.021	4.889	.027	1160.700	.604	.437
Indulgence	1	1140.461	2.321	.128	1149.178	2.206	.138

TABLE V: Optimistic Model for Perception of Quality and Enjoyment

Parameter	df_{num}	Perceived Quality			Enjoyment		
		df_{den}	F	p	df_{den}	F	p
Movie Excerpt	11	176.430	11.260	.000	179.877	46.990	.000
Frame Rate (FR)	2	1086.420	28.464	.000	1116.890	8.025	.000
Frame Dimension (Dim)	1	1100.669	17.950	.000	1120.818	3.130	.077
Bit-Rate (BR)	1	1092.200	13.052	.000	1121.960	.054	.816
FR * Dim	2	1091.110	2.892	.056	1117.780	.719	.487
FR * BR	2	1103.450	5.269	.005	1127.280	1.488	.226
Dim * BR	1	1114.040	7.513	.006	1128.860	1.466	.226
FR * Dim * BR	2	1087.310	7.143	.001	1113.480	.020	.980

HybridPlacesCNN is trained on 1.3 million images from the ImageNet challenge 2012 [33] and 2.5 million images on the Places dataset [68] on 205 scene categories. Caffe [29] was used to extract the features from the ReLu layer following fc7 layer.

Adjective-Noun Pairs (ANP) is an attribute representation based on emotion related concepts [5]. Classifiers are used to detect 2089 ANPs in each frame of the videos in our dataset. Adjectives are strongly related to one of the 8 emotion categories, as put forward by Plutchik [51], and nouns correspond to objects and scenes in the frames. Emotional connection with the media that the users form might influence their perception of quality and enjoyment.

Audio Affect-factors: To represent the affective characteristics associated with audio, a range of features are extracted. These included: musical chroma features [47], prosodic features [7] as well as low-level descriptors such as intensity, loudness, MFCC features, pitch, probability of voicing, pitch envelope, line spectral frequencies, and zero-crossing rate. OpenSmile [9] was used to extract these features from the audio signals of all videos.

System-factors: To represent the quality characteristics, bit-rate, frame-rate, frame dimension, and perceptual characteristics are used. These features describe the no-reference quality metric [43], [56] for the videos in our dataset. The following features were extracted to represent the perceptual characteristics: spatial domain natural scene statistics, temporal distortions in the video, statistical DCT features reflecting the perceptual difference between pristine and distorted videos, and motion coherence feature describing the coherence in strength and direction of local motion due to temporal distortions. We chose to use existing features rather than designing new descriptors because our aim was to investigate the role of human factors in predicting perceived quality and enjoyment, rather than to improve existing descriptors. Thus, we used the code from [56], which could predict no-reference video quality, to represent the temporal distortions in the video. For other features [43], we employed the averaging technique

Human-factors: To represent human factors, the five factors of personality, six dimensions of culture, gender, age, and nationality of the users are used.

2) *Experimental Setting:* After extracting the above-mentioned features from the videos, the dataset is divided into 50%/50% training/test sets. The scores on perceptual quality and enjoyment are normalized based on z-scores and converted to binary categories of high and low. The parameters for L-1 normalized linear SVM are tuned based on a 5-fold cross validation setting at the classification task. Results are shown on the test set.

V. RESULTS

A. Statistical Analysis

1) *Baseline Model:* An analysis of this model can be seen in Table III. It can be seen that all of the system factors and their interactions had a statistically significant effect on the

perception of quality with the movie excerpt itself making a contribution. As expected, the movie excerpt itself had the largest impact on enjoyment. However, it is interesting to note that only a small number of the system factors had a statistically significant effect on enjoyment.

2) *Extended Model:* Table IV above show an analysis of the extended model. In addition, a more comprehensive overview of the parameters in the models can be found in Tables VI and VII. It can be seen that several of personal and cultural traits are statistically significant predictors. Of particular interest are those which influence both perception of quality and overall enjoyment. These were: masculinity and conscientiousness. The regression coefficients for these parameters show that they have positive and negative impacts on overall ratings, respectively. Individualism, pragmatism, and neuroticism had, respectively, negative, negative, and positive impact on perceived quality. However, openness, power distance, and uncertainty avoidance respectively had positive, negative, and negative impacts on enjoyment respectively. The magnitudes of the effect sizes are also comparable with some system factors.

3) *Optimistic Model:* An analysis of the optimistic models is shown in Table V. There is only a small number of differences between the baseline and the optimistic model. As expected, the F-statistics for the intercepts are much larger, showing that they explain a larger proportion of the variance. Additionally, the borderline significant interaction between frame rate and frame dimension has become non-significant. The most notable difference, however, is a large increase in the variance explained as a result of including participants as random effects.

4) *Model Comparison:* The models are compared using paired t-tests on the Mean Squared Residuals (MSR), shown in Tables VIII and IX, and the proportional reduction in overall mean squared error of prediction is examined (see [6]).

Models for Perception of Quality: In the baseline model, the MSR is 1.2636 ($\sigma = 1.77$). The optimistic model reduces the MSR to 0.9085 ($\sigma = 1.63$) ($p < .000$). This represents 24.2% of the overall variance predicted (compared to 37.9% overall). However, culture and personality only predict a small proportion of this variance. The extended model predicts approximately 9.3% of variance attributable to human factors, reducing the baseline MSR to 1.2311 ($\sigma = 1.77$) ($p < .014$).

Models for Enjoyment: In the baseline model, the MSR is 1.3684 ($\sigma = 1.63$). The optimistic model reduces the MSR to 0.9481 ($\sigma = 1.22$) ($p < .000$). This represents 23.0% of the overall variance predicted (compared to 47.8% overall). However, again, culture and personality only predict a small proportion. The extended model predicts approximately 9.3% of variance attributable to human factors, reducing the baseline MSR to 1.3290 ($\sigma = 1.58$) ($p < .001$).

5) *Quality and Enjoyment:* Descriptive statistics for each movie excerpt and parameter setting are presented alongside a correlation analysis between quality and enjoyment in Figures 4,2 and Table X. These show how the parameters, the content, and level of enjoyment interact when human factors are not

TABLE VI: Standardised Parameter Estimates in the Extended Perceived Quality Model

Source	Parameter	Estimate	$SE_{\bar{x}}$	df	t	p	95% CI		$r_{y\lambda}$
							Lower	Upper	
System	5fps**	-.5568	.129	1086.514	-4.320	.000	-.810	.304	-.129
	384k**	-.8580	.134	1154.374	-6.377	.000	-1.122	-.594	-.184
	5fps * 384k**	.8026	.188	1145.181	4.268	.000	.433	1.171	.125
	15fps * 384k**	.7103	.182	1166.053	3.886	.000	.351	1.068	.113
	480p * 384k**	.8595	.191	1154.370	4.485	.000	.483	1.235	.130
	5fps * 480p * 384k**	-.8019	.262	1151.397	-3.061	.002	-1.316	-.287	-.089
	15fps * 480p * 384k**	-.7810	.260	1163.006	-3.003	.003	-1.291	-.270	-.087
Personality	Conscientiousness**	-.0747	.028	1141.817	-2.669	.008	-.130	-.020	-.078
	Neuroticism**	.0943	.028	1149.100	3.422	.001	.040	.148	.100
Culture	Individualism*	-.0636	.027	1149.721	-2.349	.019	-.117	-.010	-.069
	Masculinity*	.0659	.028	1147.422	2.362	.018	.011	.121	.069
	Pragmatism*	-.0653	.030	1152.021	-2.211	.027	-.123	-.007	-.065

Reference categories were: Frame Rate = 25fps, Frame Dimension = 720p, Bit Rate = 768k.

† $p < .10$, * $p < .05$, ** $p < .01$

TABLE VII: Standardized Parameter Estimates in the Extended Enjoyment Model

Source	Parameter	Estimate	$SE_{\bar{x}}$	df	t	p	95% CI		$r_{y\lambda}$
							Lower	Upper	
Personality	Conscientiousness*	-.0601	.026	1141.249	-2.296	.022	-.111	-.009	-.067
	Openness*	.0528	.025	1145.365	2.084	.037	-.003	.103	.061
Culture	Power Distance**	-.0795	.026	1152.465	-3.023	.003	-.131	-.028	-.088
	Masculinity†	.0474	.026	1141.312	1.820	.069	-.004	.099	.053
	Uncertainty Avoidance*	-.0661	.028	1144.106	-2.398	.017	-.120	-.012	-.070

Reference categories were: Frame Rate = 25fps, Frame Dimension = 720p, Bit Rate = 768k. All system interaction effects are non-significant and so are not shown.

† $p < .10$, * $p < .05$, ** $p < .01$

TABLE VIII: Paired t-Test Comparing Models for Perceived Quality on MSR

Models	$\Delta\bar{x}$	σ	$SE_{\bar{x}}$	95% CI		t	df	p
				Lower	Upper			
<i>Baseline</i> → <i>Extended</i>	.0325	.461	.013	.007	.058	2.472	1231	.014
<i>Baseline</i> → <i>Optimistic</i>	.3551	1.009	.029	.299	.412	12.350	1231	.000

TABLE IX: Paired t-Test Comparing Models for Enjoyment on MSR

Models	$\Delta\bar{x}$	σ	$SE_{\bar{x}}$	95% CI		t	df	p
				Lower	Upper			
<i>Baseline</i> → <i>Extended</i>	.0394	.430	.012	.015	.063	3.219	1231	.001
<i>Baseline</i> → <i>Optimistic</i>	.4199	1.129	.032	.357	.483	13.069	1231	.000

TABLE X: Correlation Analysis Showing the Relationship Between Perceived Quality and Enjoyment

	C-I	C-II	C-III	C-IV	C-V	C-VI	C-VII	C-VIII	C-IX	C-X	C-XI	C-XII	Total ^b
Spearman's ρ	.252	.170	.377	.161	.242	.447	.507	.439	.367	.269	.391	.369	.375
p	.007	.082	.000	.095	.011	.000	.000	.000	.000	.007	-.000	.000	.000

^bThis aggregate is based on non-independent data due to repeated measures.

Movie Excerpts: C-I: A_FISH_CALLED_WANDA; C-II: AMERICAN_HISTORY_X; C-III: CHILDS_PLAY_JI; C-IV: COPYCAT; C-V: DEAD_POETS_SOCIETY_1;

C-VI: DEAD_POETS_SOCIETY_2; C-VII: FOREST_GUMP; C-VIII: SE7EN_1; C-IX: SE7EN_3; C-X: SOMETHING_ABOUT_MARY; C-XI: THE_PROFESSIONAL; C-XII: TRAINSPOTTING.

controlled. It can be seen that the overall correlation between quality and enjoyment is significant, however this is not consistent across all of the movie excerpts. Additionally, the 'highest' quality parameters do not consistently perform well.

B. Low System Parameters

The left part of Figure 5 shows ratings on enjoyment for different clips with low system parameter settings. For three clips (namely DEAD_POETS_SOCIETY_2, FOREST_GUMP and SE7EN_1), there were a higher (or equal) percentage of ratings saying that users enjoyed these clips (rating >3) than those who said that they didn't enjoy

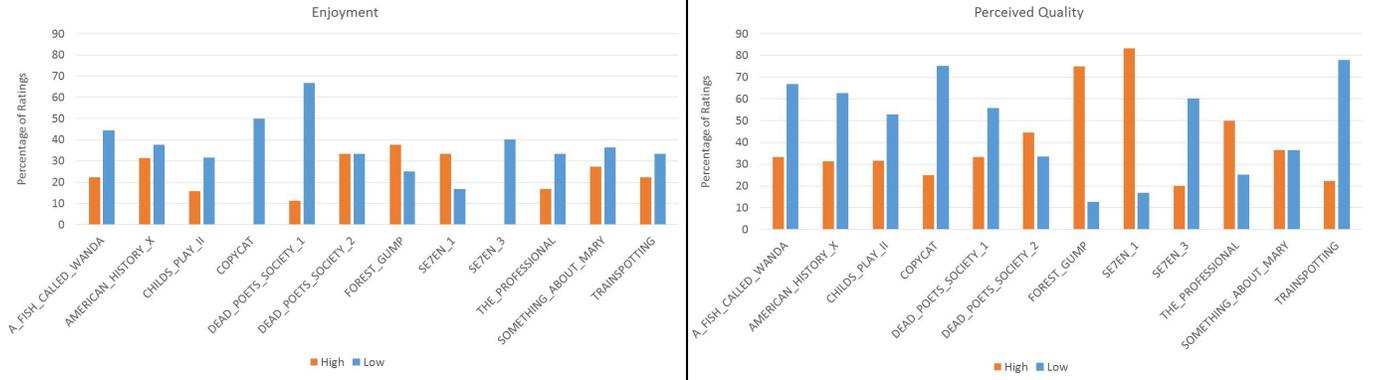


Fig. 5: Distribution of ratings on Low system parameter setting clips.

(rating <3).

The right part of Figure 5 shows ratings on perceived quality for different clips with low system parameter settings. As with ratings on enjoyment, a similar trend can be seen in perceived quality ratings for `DEAD_POETS_SOCIETY_2`, `FOREST_GUMP` and `SE7EN_1`. Additionally, `THE_PROFESSIONAL` and `SOMETHING_ABOUT_MARY` also got higher percentage of high ratings on perceived quality.

To explain the above rating behavior, consider the example of `FOREST_GUMP`. This has a very unique sequence where the protagonist sees his son for the first time, intensely affecting the audience. Such affective responses are based on a strong interplay between both human factors and nature of the content, perhaps overpowering the influence of system parameter settings.

Analyzing those clips which received higher ratings despite having lower system parameters revealed that agreeableness had a borderline significant influence on ratings on enjoyment ($p = 0.08$) and a significant influence on perceived quality ($p = 0.02$). This finding aligns with work suggesting that agreeable people have higher tolerances [42].

C. Predictive Models

The accuracy of the trained models at predicting perceived quality and enjoyment is shown in Table XI. Four different sets of factors (namely human, system, content and emotion) were considered and appropriate features were extracted to represent the same as described above.

For perceived quality, the model trained on emotion factors outperforms others. In [52], web videos with particular emotions significantly varied the MOS scores that users gave on perceived quality. Similar observations were made in mobile videos [30] and also in studies where aesthetics were seen to be significant in predicting video quality [45]. Aesthetics factors were shown to contribute significantly in modeling emotions in previous works [39].

Again, extending the system, content and emotion factors using the additional human factors revealed that combining human and system factors provides the best performance. However, they do not change the performance of the model trained on content factors. They in fact reduce the performance

TABLE XI: Accuracy of Predictive Models using different features

Feature	Perceived Quality	Enjoyment
Human (H)	64.29	56.82
System (S)	64.29	56.82
Content (C)	63.15	69.81
Emotion (E)	65.58	69.81
H+S	66.07	65.91
H+C	63.15	69.81
H+E	64.45	70.45
H+S+C	62.82	67.53
H+S+E	66.07	68.99
H+C+E	64.77	69.64
H+S+C+E	62.66	65.91

of the model trained on emotion factors. This could be due to the problem of dimensionality [61]. Note that this is the test set performance and not training/cross-validation set performance. A similar issue was seen in the models trained by taking three (and then all four) factors at once as input.

For enjoyment, the models trained on content and emotion factors show better performance than others. This was partially seen even in our statistical analysis where the content was significantly correlated with enjoyment. The significance of emotional factors was studied elsewhere [2], [4], [62]. When human factors are combined with other factors, the model trained on human and emotion factors outperform the others, giving the best performance in predicting enjoyment of videos.

These results show that factors other than perceptual characteristics (which are used in previous works on modeling quality of experience [69], [70]) should be considered to build a more comprehensive model.

VI. DISCUSSION

These results show that human factors play a key role in the way perception of quality and enjoyment are rated. The analysis of perceptual quality, in particular, indicated that a greater proportion of the variance can be predicted by human factors (24.3%) than by system factors (13.7%); however,

all the system factors and most of their interactions have larger effect sizes than any individual human factor. This implies that perceived quality and enjoyment are determined by humans as much as they are determined by multimedia systems. This follows prior findings in the area [14], but more importantly demonstrates that “lower” system factors may not automatically entail lower quality or enjoyment. As an example, the parameter setting {25fps, 480p, 384k} was ranked 4th for perceived quality and 1st for enjoyment, despite having a low bit-rate. Indeed, performance varied across different movie excerpts and participants. As such, understanding these factors could be used to prioritize limited resources while maintaining acceptable quality.

The human factors explored in this study, namely personality and culture, represent a small portion of the variance which can be attributed to human factors. Collectively, both sets of variables represent 9.3% of the variance. While this is a non-trivial proportion, further study is needed to discover other contributing factors. In this regard, a key facet to consider is moderation. That is, where the magnitude (and sign) of a relationship (e.g., between perceived quality and enjoyment) depends on the value of a third variable (e.g., personality). A correlation analysis of the relationship between perceived quality and enjoyment shows considerable inconsistency in effect size across different movie excerpts. In particular, excerpts with objectionable content (i.e., graphic murders) were not significant, while excerpts with widely acceptable content (i.e., romance) had large correlations. Presumably, this is because people do not enjoy objectionable experiences and so quality is not an important factor in such cases. As viewers tend to object to different content, such interactions could be used as a basis for managing quality of service parameters.

Due to the international nature of the research presented in this article, participants did not use the same laboratories and therefore did not use the same devices. As such, system and contextual factors such as dead pixels, lighting conditions, and differences in computer hardware could have confounded any effect that has otherwise been attributed to the participants themselves. Consequently, this may over-estimate the variance attributable to human factors and so the relative 9.3% contribution of personality and culture may be an under-estimation.

It should be noted that the indulgence trait did not predict either perceived quality or enjoyment. This is interesting because habituation and sensitization effects were anticipated. That is, those with high indulgence scores tend to concentrate on individual well being and leisure time. It follows, then, that they may seek to immerse themselves in multimedia content to a greater extent than those with lower indulgence scores. It is possible, however, that the indulgence may not correspond directly with multimedia use, specifically. Whether or not participants use multimedia services may, therefore, be an important factor to consider in future studies.

Prior research [63] suggests that agreeableness is a predictor for perceptual quality whereas extroversion is a predictor for enjoyment. Other studies suggest that there were no significant influence of personality in perceived quality [69]. Such

differences could correspond to: stimuli-oriented interaction effects (YouTube videos [69], IPTV [63] vs. affective movie excerpts); use of different measurement instruments (TIPI for personality in [69] vs. BFI-10 in this study); variation in samples, sample sizes, and sampling method (single institution [63], [69] vs. several universities in different countries); analysis technique used (linear classifiers in [69] vs. statistical modeling); and so on. To address such differences, building a comprehensive QoE model involving data from multiple datasets is encouraged.

VII. CONCLUSION

This study presents further evidence that multimedia perception of quality and enjoyment are influenced by an intricate interplay between system, context, and human factors. Further to this, the model proposed indicates that human factors play a critical role, with key findings as follows:

- Approximately 13.7% of the variance in perceived quality is predicted by system factors while 24.3% is predicted by human factors.
- Around 24.8% of the variance in enjoyment can predicted by system factors while 23.1% can be predicted by human factors.
- About 9.3% of the variance attributable to human factors can be predicted by personality and cultural traits (for both perceived quality and enjoyment).
- The traits of masculinity and conscientiousness are important predictors for both perceived quality and enjoyment.
- Individualism, pragmatism, and neuroticism are important predictors for perceived quality.
- Power distance, uncertainty avoidance, and openness are important predictors for enjoyment.
- The quality-enjoyment correlation varies in magnitude across the movie excerpts.
- Predictive model trained on a combination of human factors and system factors give best performance of predicting perceived quality.
- Predictive model trained on a combination of human, content and emotion factors represent best performance of predicting enjoyment.

Perceptual quality and enjoyment are as much human constructs as they are the result of objective technological differences. As such, analysis of these factors can help system designers to optimize perceived quality and enjoyment under conditions where content delivery is constrained. Based on the results of this study, a human-centered quality of service algorithm could incorporate individualism, masculinity, pragmatism, neuroticism, and conscientiousness as parameters, based upon their ability to predict perceived quality; and they may even account for individual preference for different content, based upon the inconsistent correlations between quality and enjoyment.

Further modeling of human factors is needed because the personality and cultural traits selected in this study only represented a small proportion of the variance. In addition, the following three limitations exist. Firstly, only the main

linear effects of personality and culture were explored. It is possible that these traits interact with other factors and so more complex relationships may exist. Second, the scope of the study was limited to a small subset of human and system factors. Broader investigations of system, context, and human factors are needed. Third, those recruited in this study were university students. Although there is no evidence to suggest that students are different to the general population in terms of quality of experience, additional work is warranted to ensure the full range of each human factor is considered (see [23], [31]) and any potential confounds are identified.

REFERENCES

- [1] A. Acar, T. Taura, E. Yamamoto, and N. Yusof. Object vs. relation: understanding the link between culture and cognition with the help of wordnet. *International Journal on Asian Language Processing*, 21(9):199–208, 2011.
- [2] S Bardzell, J Bardzell, and T Pace. Understanding affective interaction: Emotion, engagement, and internet videos. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–8. IEEE, 2009.
- [3] S Bhattacharya, B Nojavanasghari, T Chen, D Liu, SF Chang, and M Shah. Towards a comprehensive computational model for aesthetic assessment of videos. In *21st ACM Multimedia*, pages 361–364. ACM, 2013.
- [4] H Bilandzic and RW Busselle. Enjoyment of films as a function of narrative experience, perceived realism and transportability. *Communications*, 36(1):29–50, 2011.
- [5] D Borth, T Chen, R Ji, and SF Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *21st ACM international conference on Multimedia*, 2013.
- [6] R. Bosker and T. Snijders. Multilevel analysis: An introduction to basic and advanced multilevel modeling, 2nd ed. *New York*, 2012.
- [7] MJ Carey, ES Parris, H Lloyd-Thomas, and S Bennett. Robust prosodic features for speaker identification. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1800–1803. IEEE, 1996.
- [8] M Cristani, A Vinciarelli, C Segalin, and A Perina. Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In *21st ACM international conference on Multimedia*. ACM, 2013.
- [9] F Eyben, M Woellmer, and B Schuller. the munich open speech and music interpretation by large space extraction toolkit. 2010.
- [10] M. Fiedler, T. Hossfeld, and P. Tran-Gia. A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE*, 24(2):36–41, 2010.
- [11] X.P. Gao, J.H. Xin, T. Sato, A. Hansuebsai, M. Scalzo, K. Kajiwaru, S.S. Guan, J. Valdeperas, M.J Lis, and M. Billger. Analysis of cross-cultural color emotion. *Color Research & Application*, 32(3):223–229, 2007.
- [12] G. Ghinea and S.Y Chen. The impact of cognitive styles on perceptual distributed multimedia quality. *British Journal of Educational Technology*, 34(4):393–406, 2003.
- [13] G Ghinea and SY Chen. Perceived quality of multimedia educational content: A cognitive style approach. *Multimedia systems*, 11(3):271–279, 2006.
- [14] G. Ghinea and J.P Thomas. Qos impact on user perception and understanding of multimedia video clips. In *Proceedings of the sixth ACM international conference on Multimedia*, pages 49–54. ACM, 1998.
- [15] L.R Goldberg. An alternative” description of personality”: the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.
- [16] E. Goldstein. *Sensation and Perception*. Cengage Learning, 2013.
- [17] E.B. Goldstein. *Cultural effects on visual perception*. SAGE Publications, Inc., Encyclopedia of Perception, 2010.
- [18] S.D. Gosling, P.J. Rentfrow, and W.B. Swann. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.
- [19] Video Quality Experts Group et al. Report on the validation of video quality models for high definition video content. *Tech. Report*, <http://www.vqeg.org>, 2010.
- [20] S.R. Gulliver and G. Ghinea. Defining user perception of distributed multimedia quality. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(4):241–257, 2006.
- [21] S.C. Guntuku, S. Roy, and W. Lin. Personality modeling based image recommendation. In *MultiMedia Modeling*, pages 171–182. Springer, 2015.
- [22] S.C. Guntuku, M.J. Scott, H. Yang, G. Ghinea, and W. Lin. The CP-QAE-I: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia. In *QoMEX*. IEEE, 2015.
- [23] David S Hands. A basic multimedia quality model. *Multimedia, IEEE Transactions on*, 6(6):806–816, 2004.
- [24] G. Hofstede. Dimensionalizing cultures: The hofstede model in context. *Online readings in psychology and culture*, 2(1):8, 2011.
- [25] G. Hofstede, G. J. Hofstede, M. Minkov, and H. Vincken. Values survey module 2013. URL: <http://www.geerthofstede.nl/vsm2013>, 2013.
- [26] G. Hofstede, G.J. Hofstede, and M. Minkov. *Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival*. McGraw-Hill, 2010.
- [27] Tobias Hofffeld, Christian Keimel, Matthias Hirth, Bruno Gardlo, Julian Habigt, Klaus Diepold, and Phuoc Tran-Gia. Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing. *Multimedia, IEEE Transactions on*, 16(2):541–558, 2014.
- [28] M. Hyder, N. Crespi, M. Haun, and C. Hoene. Are qoe requirements for multimedia services different for men and women? analysis of gender differences in forming qoe in virtual acoustic environments. In *Emerging Trends and Applications in Information Communication Technologies*, pages 200–209. Springer, 2012.
- [29] Y Jia, E Shelhamer, J Donahue, S Karayev, J Long, R Girshick, S Guadarrama, and T Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [30] S Jumisko-Pyykkö and J Häkkinen. Evaluation of subjective video quality of mobile devices. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 535–538. ACM, 2005.
- [31] Asiya Khan, Lingfen Sun, and Emmanuel Ifeachor. Qoe prediction model and its application in video quality adaptation over umts networks. *Multimedia, IEEE Transactions on*, 14(2):431–442, 2012.
- [32] Taewan Kim, Jiwoo Kang, Sanghoon Lee, and Alan C Bovik. Multimodal interactive continuous scoring of subjective 3d video quality of experience. *Multimedia, IEEE Transactions on*, 16(2):387–402, 2014.
- [33] A Krizhevsky, I Sutskever, and GE Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [34] Jong-Seok Lee. On designing paired comparison experiments for subjective multimedia quality assessment. *Multimedia, IEEE Transactions on*, 16(2):564–571, 2014.
- [35] Jong-Seok Lee, Francesca De Simone, and Touradj Ebrahimi. Subjective quality evaluation via paired comparison: application to scalable video coding. *Multimedia, IEEE Transactions on*, 13(5):882–893, 2011.
- [36] Sanghoon Lee, Marios S Pattichis, and Alan C Bovik. Foveated video quality assessment. *Multimedia, IEEE Transactions on*, 4(1):129–132, 2002.
- [37] M. Leszczuk, L. Janowski, and M. Barkowsky. Freely available large-scale video quality assessment database in full-hd resolution with h.264 coding. In *IEEE Globecom 2013*, volume -, pages pp.1–6. Atlanta, United States, Dec 2013.
- [38] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *In Eur. Conf. Comp. Vis. (ECCV)*, pages 386–399, 2008.
- [39] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM, 2010.
- [40] B. Manav. Color-emotion associations and color preferences: A case study for residences. *Color Research & Application*, 32(2):144–150, 2007.
- [41] W.A. Mansilla, A. Perkis, and T. Ebrahimi. Implicit experiences as a determinant of perceptual quality and aesthetic appreciation. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 153–162. ACM, 2011.
- [42] G. Matthews, I.J Deary, and M.C. Whiteman. *Personality traits*. Cambridge University Press, 2003.
- [43] A Mittal, AK Moorthy, and AC Bovik. No-reference image quality assessment in the spatial domain. *Image Processing, IEEE Transactions on*, 21(12):4695–4708, 2012.
- [44] Y. Miyamoto. *Culture and analytic versus holistic cognition: toward multilevel analyses of culture influences*. Academic Press, Advances in Experimental Social Psychology, 2013.

- [45] AK Moorthy, P Obrador, and N Oliver. Towards computational models of the visual aesthetic appeal of consumer videos. In *Computer Vision—ECCV 2010*, pages 1–14. Springer, 2010.
- [46] A.K. Moorthy, P. Obrador, and N. Oliver. Towards computational models of visual aesthetic appeal of consumer videos. In *In Eur. Conf. Comp. Vis. (ECCV)*, pages 1–14, 2010.
- [47] M Müller and S Ewert. Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), 2011. hal-00727791, version 2-22 Oct 2012*. Citeseer, 2011.
- [48] Matteo Naccari, Marco Tagliasacchi, and Stefano Tubaro. No-reference video quality monitoring for h. 264/avc coded video. *Multimedia, IEEE Transactions on*, 11(5):932–946, 2009.
- [49] R.E. Nisbett and Y. Miyamoto. The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10):467–473, Oct 2005.
- [50] P. Paudyal, F. Battisti, and M. Carli. A study on the effects of quality of service parameters on perceived video quality. In *Procs. of 5th European Workshop on Visual Information Processing, EUVIP*, 2014.
- [51] R Plutchik. *Emotion: A psychoevolutionary synthesis*. Harper & Row New York, 1980.
- [52] B. Rainer, M. Waltl, E. Cheng, M. Shujau, C. Timmerer, S. Davis, I. Burnett, C. Ritz, and H. Hellwagner. Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pages 278–283. IEEE, 2012.
- [53] J.A. Redi, Y. Zhu, H. de Ridder, and I. Heynderickx. How passive image viewers became active multimedia users. In *Visual Signal Quality Assessment*, pages 31–72. Springer, 2015.
- [54] Amy R Reibman, Vinay Vaishampayan, Yegnaswamy Sermadevi, et al. Quality monitoring of video over a packet network. *Multimedia, IEEE Transactions on*, 6(2):327–334, 2004.
- [55] U. Reiter, K. Brunnstrm, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank. *Factors Influencing Quality of Experience*. Springer, Quality of Experience, 2014.
- [56] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind prediction of natural video quality. *Image Processing, IEEE Transactions on*, 23(3):1352–1365, 2014.
- [57] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 24(7):1153–1172, 2010.
- [58] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, and L.K. Cormack. Study of subjective and objective quality assessment of video. *Image Processing, IEEE Transactions on*, 19(6):1427–1441, 2010.
- [59] E. Siahaan, J.A. Redi, and A. Hanjalic. Beauty is in the scale of the beholder: Comparison of methodologies for the subjective assessment of image aesthetic appeal. In *QoMEX*, pages 245–250, Sept 2014.
- [60] Wei Song and Dian W Tjondronegoro. Acceptability-based qoe models for mobile video. *Multimedia, IEEE Transactions on*, 16(3):738–750, 2014.
- [61] GV Trunk. A problem of dimensionality: A simple example. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (3):306–307, 1979.
- [62] VT Visch, ES Tan, and D Molenaar. The emotional and cognitive effect of immersion in film viewing. *Cognition and Emotion*, 24(8):1439–1445, 2010.
- [63] I. Wechsung, M. Schulz, K.P. Engelbrecht, J. Niemann, and S. Möller. All users are (not) equal—the influence of user characteristics on perceived quality, modality choice and performance. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 175–186. Springer, 2011.
- [64] Stefan Winkler and Christof Faller. Perceived audiovisual quality of low-bitrate multimedia content. *Multimedia, IEEE Transactions on*, 8(5):973–980, 2006.
- [65] Qianqian Xu, Jiechao Xiong, Qingming Huang, and Yuan Yao. Online hodgerank on random graphs for crowdsorceable qoe evaluation. *Multimedia, IEEE Transactions on*, 16(2):373–386, 2014.
- [66] J Yang, YG Jiang, AG Hauptmann, and CW Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 197–206. ACM, 2007.
- [67] Guangtao Zhai, Jianfei Cai, Weisi Lin, Xiaokang Yang, Wenjun Zhang, and Minoru Etoh. Cross-dimensional perceptual quality assessment for low bit-rate videos. *Multimedia, IEEE Transactions on*, 10(7):1316–1324, 2008.
- [68] B Zhou, A Lapedriza, J Xiao, A Torralba, and A Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.
- [69] Y Zhu, I Heynderickx, A Hanjalic, and JA Redi. Towards a comprehensive model for predicting the quality of individual visual experience. In *IS&T/SPIE Electronic Imaging*, pages 93940A–93940A. International Society for Optics and Photonics, 2015.
- [70] Y Zhu, I Heynderickx, and JA Redi. Understanding the role of social context and user factors in video quality of experience. *Computers in Human Behavior*, 49:412–426, 2015.