

# News from the AVT Department

## Archive of reports 2017 until 1st half of 2021

(in chronological reverse order)

### Modular Framework and Instances of Pixel-based Video Quality Models for UHD-1/4K

#### New Publication

Steve Göring, Rakesh Rao Ramachandra Rao, Bernhard Feiten, and Alexander Raake

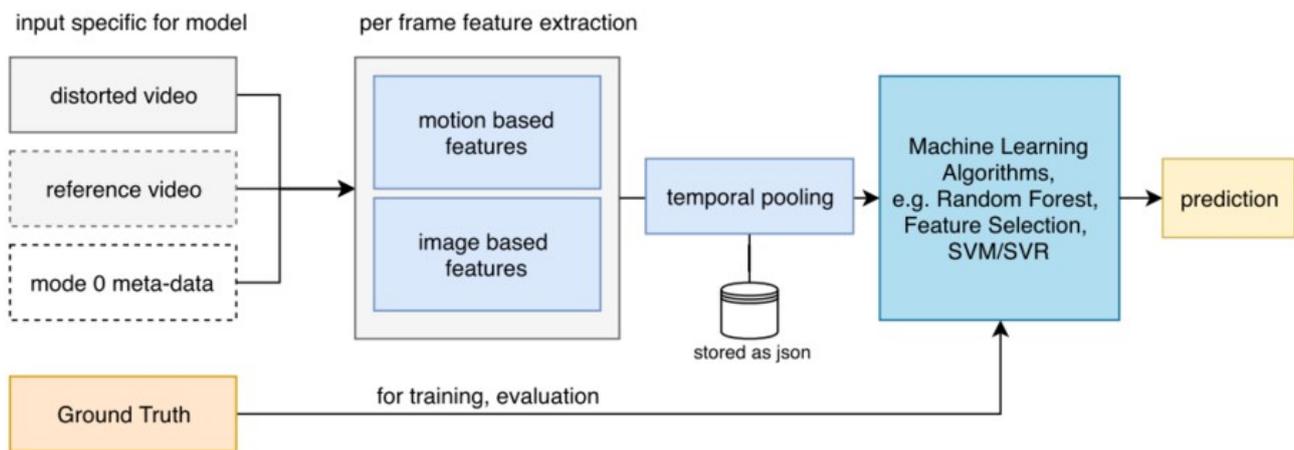
IEEE Access. vol. 9. 2021. (<https://ieeexplore.ieee.org/document/9355144>)

#### Abstract

The popularity of video on-demand streaming services increased tremendously over the last years. Most services use http-based adaptive video streaming methods. Today's movies and TV shows are typically

recorded in UHD-1/4K and streamed using settings attuned to the end-device and current network conditions. Video quality prediction models can be used to perform an extensive analysis of video codec settings to ensure high quality. Hence, we present a framework for the development of pixel-based video quality models. We instantiate four different model variants (hyfr, hyfu, fume and nofu) for short-term video quality estimation targeting various use cases. Our models range from a no-reference video quality model to a full-reference model including hybrid model extensions that incorporate client accessible meta-data.

All models share a similar architecture and the same core features, depending on their mode of operation. Besides traditional mean opinion score prediction, we tackle quality estimation as a classification and multi-output regression problem. Our performance evaluation is based on the publicly available AVT-VQDB-UHD-1 dataset. We further evaluate the introduced center-cropping approach to speed up calculations. Our analysis shows that our hybrid full-reference model (hyfr) performs best, e.g. 0.92 PCC for MOS prediction, followed by the hybrid no-reference model (hyfu), full-reference model (fume) and no-reference model (nofu). We further show that our models outperform popular state-of-the-art models. The introduced features and machine-learning pipeline are publicly available for use by the community for further research and extension.



## DFG-Project ECoClass-VR started

Recently, the Deutsche Forschungsgemeinschaft (DFG) accepted a submitted project proposal within the DFG priority programme "Auditory Cognition in Interactive Virtual Environments" ([AUDICTIVE](#)). The project is being carried out in cooperation with the Chair for Hearing Technology and Acoustics (Prof. Janina Fels, RWTH Aachen) and the Department of Cognitive and Developmental Psychology (Prof. Maria Klatte, TU Kaiserslautern).

### Project description

In the project "Evaluating cognitive performance in classroom scenarios using audiovisual virtual reality" (ECoClass-VR) we are investigating the suitability of audio-visual Immersive Virtual Environments (IVEs) for a "real-world" assessment of the influence of the visuospatial and acoustic environment on cognitive performance of adults and children in classroom-type environments. Existing knowledge on the influence of environmental variables on cognitive performance in classrooms comes predominantly from auditory experimental paradigms with typically simple acoustic replications. So far, only limited attention has been paid to visual processing, without considering relevant audiovisual aspects.

Read more in the [project page](#)!

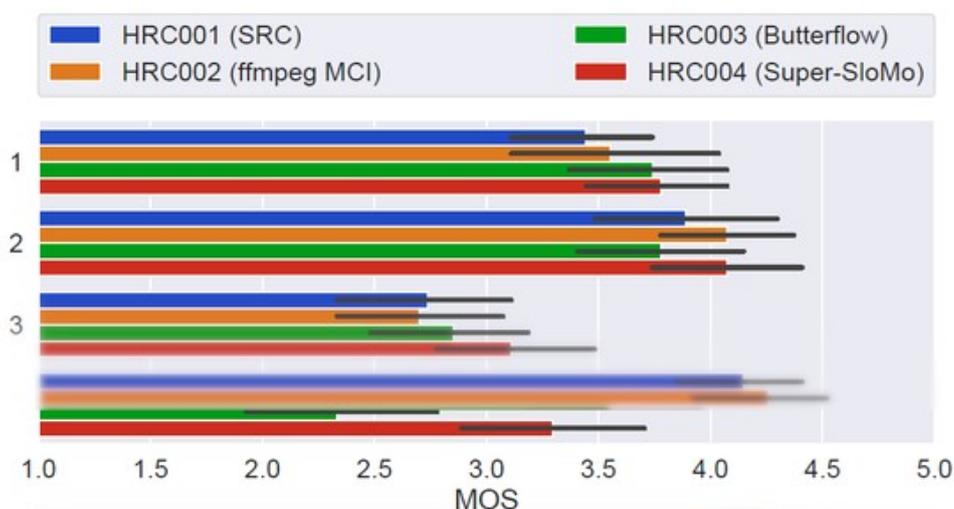
## Between the Frames - Evaluation of Various Motion Interpolation Algorithms to Improve 360° Video Quality

22nd IEEE International Symposium on Multimedia. Online. December 2020

**Stephan Fremerey, Frank Hofmeyer, Steve Göring, Dominik Keller, Alexander Raake**

With the increasing availability of 360° video content, it becomes important to provide smoothly playing videos of high quality for end users. For this reason, we compare the influence of different Motion Interpolation (MI) algorithms on 360° video quality. After conducting a pre-test with 12 video experts, we found that MI is a useful tool to increase the QoE (Quality of Experience) of omnidirectional videos. As a result of the pretest, we selected three suitable MI algorithms, namely ffm-

peg Motion Compensated Interpolation (MCI), Butterflow and Super-SloMo. Subsequently, we interpolated 15 entertaining and realworld omnidirectional videos with a duration of 20 seconds from 30 fps (original framerate) to 90 fps, which is the native refresh rate of the HMD used, the HTC Vive Pro. To assess QoE, we conducted two subjective tests with 24 and 27 participants. In the first test we used a Modified Paired Comparison (M-PC) method, and in the second test the Absolute Category Rating (ACR) approach. In the M-PC test, 45 stimuli were used and in the ACR test 60. Results show that for most of the 360° videos, the interpolated versions obtained significantly higher quality scores than the lower-framerate source videos, validating our hypothesis that motion interpolation can improve the overall video quality for 360° video. As expected, it was observed that the relative comparisons in the M-PC test result in larger differences in terms of quality. Generally, the ACR method lead to similar results, while reflecting a more realistic viewing situation. In addition, we compared the different MI algorithms and can conclude that with sufficient available computing power Super-SloMo should be preferred for interpolation of omnidirectional videos, while MCI also shows a good performance.



## AVT lab nominated for the Thuringian Innovation Award



The awards were presented on November 25, 2020 in Weimar. The prize in the category "Digital & Media" was finally awarded to **room AG Jena** for "EXPO-X, the platform for virtual and hybrid events".

[Website of the Innovation Award](#) (with all nominees and award winners)



The awards were presented on 25 November 2020 in Weimar.

The prize in the category "Digital & Media" was finally awarded to room AG Jena for "EXPO-X, the platform for virtual and hybrid events".

[Website of the Innovation Award](#) (with all nominees and award winners)

## CO-HUMANICS project adopted

University press release: [Mit Augmented Reality und Robotik soziale Kontakte älterer Menschen stärken](#)

The Carl-Zeiss-Stiftung will support the CO-HUMANICS project ("Co-Presence of Humans and Interactive Companions for Seniors") of the TU Ilmenau within the call "Durchbrüche" over the next five years.

The following groups are involved in this project:

- Audiovisuelle Technik (AVT), Prof. Raake, Fak. EI (Speaker)
- Elektronische Medientechnik (EMT), Prof. Brandenburg/Dr. Werner and successor, Fak. EI
- Virtuelle Welten und Digitale Spiele (VWDS), Prof. Broll, Fak. WM
- Medienpsychologie und Medienkonzeption (MPMK), Prof. Döring, Fak. WM
- Neuroinformatik und Kognitive Robotik (NEUROB), Prof. Groß, Fak. IA

In addition, the University Computer Centre (UniRZ), the Virtual Reality Competence Centre (KVR) and the Fraunhofer IDMT will contribute to the realisation.

Contribution by Mitteldeutscher Rundfunk as part of the "mdr Wissen" series: [Mit Augmented Reality gegen Einsamkeit im Alter](#)

### University press release

#### [TU Ilmenau: Using Augmented Reality and Robotics to Strengthen Social Contacts of Elderly People](#)

In a recently approved research project, the Technical University of Ilmenau will develop state-of-the-art methods to enable elderly people to have contact with familiar people from far away in their home environment. These people - relatives or friends, but also medical or nursing staff - will

be "projected" into the environment using novel technical methods as if they were there themselves. The project is being funded by the Carl-Zeiss-Stiftung with up to 4.5 million euros over the next five years.

Especially senior citizens often find it difficult to maintain regular social contacts. After retirement, contacts with colleagues break off, children and relatives live far away, and for health reasons it becomes more difficult to visit fellow people. In addition, the corona pandemic is currently also restricting the opportunities for older people to maintain contact with relatives and friends. In the research project CO-HUMANICS ("Co-Presence of Humans and Interactive Companions for Seniors"), an interdisciplinary research team from Technische Universität Ilmenau and Thüringer Zentrums für Lernende Systeme und Robotik (Thuringian Center for Learning Systems and Robotics) will develop technical solutions from 2021 on to integrate remote people into the home environment as if they were there in person. Alexander Raake, Professor for Audiovisual Technology at TU Ilmenau, is the representative of the research team.

Based on augmented reality technologies, the team is developing innovative communication channels with which the "connected" subjects appear much more present to the elderly as realistic representations than is the case with conventional telephone or video calls. For example, they can reach out to their conversation partners if they are presented in an appropriate spatial manner. The so-called co-presence is not intended to replace interpersonal contact, but to promote and improve existing ties to familiar people over distances.

In addition to the new augmented reality technologies, the CO-HUMANICS project also wants to develop technical assistance systems for elderly people, which can help them to communicate with relatives or even medical care personnel at a distance and to receive assistance. These robots would be able to provide concrete help, for instance, in operating technical devices or to put a conversation partner in an optimal position for talking to the senior citizen. With the help of both systems - augmented reality and robotics - elderly people can communicate with people close to them in their home environment and receive support in everyday activities.

The CO-HUMANICS project is supported by the Carl-Zeiss-Stiftung within the framework of the "Durchbrüche 2020" ("Breakthroughs 2020") funding program with up to 4.5 million euros over the next five years. The Carl-Zeiss-Stiftung has set itself the goal of creating scope for scientific breakthroughs. As a partner of excellent science, it supports both basic research and application-oriented research and teaching in the STEM- disciplines (Science Technology Engineering Mathematics). Founded in 1889 by the physicist and mathematician Ernst Abbe, the Carl-Zeiss-Stiftung is one of the oldest and largest private science-promoting foundations in Germany. It is the sole owner of Carl Zeiss AG and SCHOTT AG. Its projects are financed from the dividend distributions of the two foundation companies.

**Contact us:**

Professor Alexander Raake  
Head of the Audiovisual Technology Group  
Phone: +49 3677 69-1468  
[alexander.raake@tu-ilmenau.de](mailto:alexander.raake@tu-ilmenau.de)

## DFG-Project SoPhoAppeal started

Recently the German Research Foundation (DFG) gave a positive assessment to a submitted project proposal about photo appeal and aesthetics.

### Project description

The project SoPhoAppeal covers topics regarding image appeal and liking. Starting with the development of an image dataset including likes, views and other social signals, crowd- and lab tests will be performed to analyze the connection of liking and aesthetic rating. Furthermore models to predict such aesthetic ratings will be developed.

Read more in the [project overview](#).

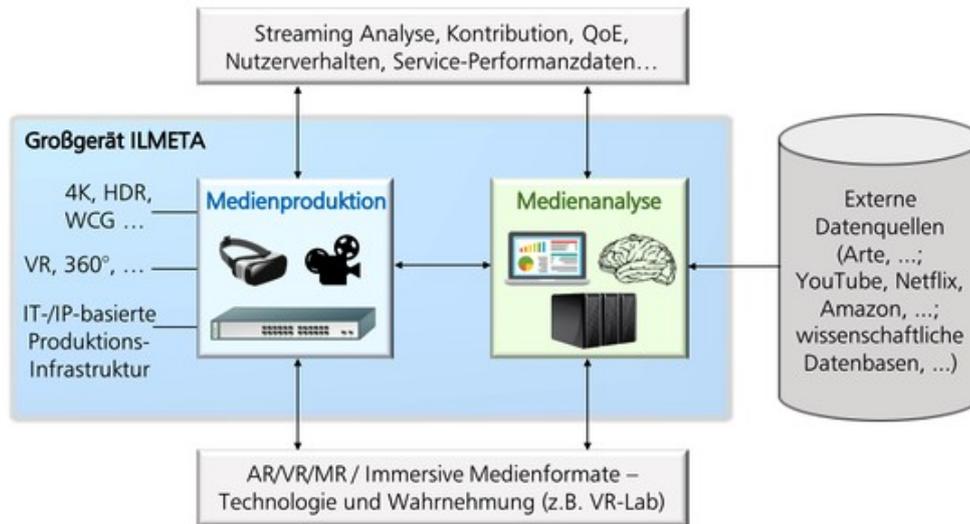
## Realization of the large-scale device ILMETA started

In the department, the implementation of the large-scale device ILMETA has been started, in which the first orders are triggered.

At the beginning of 2020, the German Research Foundation (DFG) gave a positive assessment of the department's application for the co-financing of the large-scale ILMETA facility. As a result, the DFG and the Free State of Thuringia are providing a total of €570,000 for the realisation of this infrastructure project.

### Project description

ILMETA (Interconnected Lab for MEdia Technology Analytics) is a networked large-scale device for the investigation of audiovisual media technology systems. It is intended to enable research into the measurement and evaluation of systems for recording, signal processing, network technology (production, streaming) and reproduction, by means of data analysis (signals, meta/measurement data) and tests on human perception and quality of experience (QoE). Since all systems along the end-to-end chain aim at the best possible processing of media content (content: signals and meta data), content represents the essential measurement information. For various reasons (including copyright), high-quality content that is representative for the application is only available to a limited extent for research, especially for new immersive media formats such as 360° or high-resolution video. In order to be able to achieve the research objectives, the large-scale equipment consists of two closely interlinked components: (1) production infrastructure for the creation of high-quality research content, including current IP-based studio and measurement technology, (2) computing and storage infrastructure for the analysis of the resulting heterogeneous research data (signals, measurement/metadata) in order to gain insights for the improvement of media systems for different users.



## AVT members win DASH Industry Forum Excellence Award in collaboration with TU Berlin, NTNU and TU Munich

This year's DASH Industry Forum Excellence in DASH Awards were presented at ACM MMSys 2020. The prizes were awarded for "practical enhancements and developments which can sustain future commercial usefulness of DASH". The paper "Comparing Fixed and Variable Segment Durations for Adaptive Video Streaming – A Holistic Analysis" was written by Susanna Schwarzmann (TU Berlin), Nick Hainke (TU Berlin), Thomas Zinner (NTNU Norway), and Christian Sieber (TU Munich) together with Werner Robitza and Alexander Raake from the AVT group. The paper won the first prize in the ceremony.

More info about the awards can be found here

(<https://multimediacommunication.blogspot.com/2020/06/dash-if-awarded-excellence-in-dash.html>). The paper is available here (<https://dl.acm.org/doi/abs/10.1145/3339825.3391858>).



# Bitstream-based Model Standard for 4K/UHD: ITU-T P.1204.3 – Model Details, Evaluation, Analysis and Open Source Implementation

Twelfth International Conference on Quality of Multimedia Experience (QoMEX). Athlone, Ireland. May 2020

**Rakesh Rao Ramachandra Rao, Steve Göring, Werner Robitza, Alexander Raake, Bernhard Feiten, Peter List, and Ulf Wüstenhagen**

With the increasing requirement of users to view high-quality videos with a constrained bandwidth, typically realized using HTTP-based adaptive streaming, it becomes more and more important to determine the quality of the encoded videos accurately, to assess and possibly optimize the overall streaming quality.

In this paper, we describe a bitstream-based no-reference video quality model developed as part of the latest model-development competition conducted by ITU-T Study Group 12 and the Video Quality Experts Group (VQEG), "P.NATS Phase 2". It is now part of the new P.1204 series of Recommendations as P.1204.3.

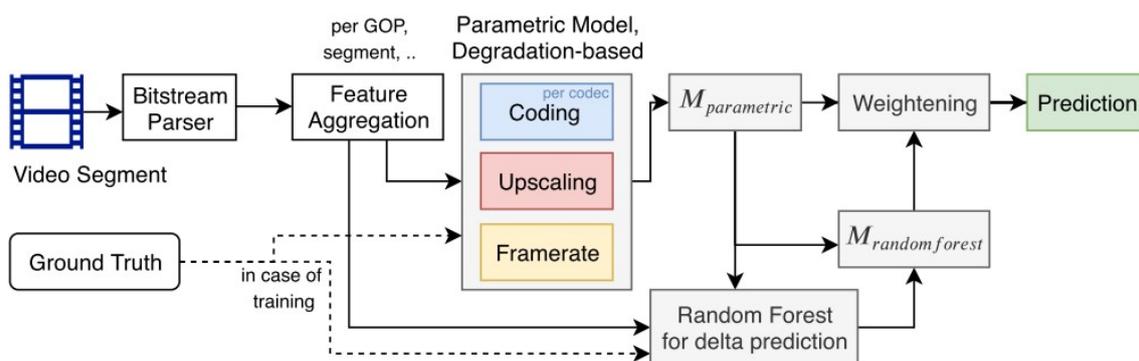
It can be applied to bitstreams encoded with H.264/AVC, HEVC and VP9, using various encoding options, including resolution, bitrate, framerate and typical encoder settings such as number of passes, rate control variants and speeds.

The proposed model follows an ensemble-modelling--inspired approach with weighted parametric and machine-learning parts to efficiently leverage the performance of both approaches. The paper provides details about the general approach to modelling, the features used and the final feature aggregation.

The model creates per-segment and per-second video quality scores on the 5-point Absolute Category Rating scale, and is applicable to segments of 5--10 seconds duration.

It covers both PC/TV and mobile/tablet viewing scenarios. We outline the databases on which the model was trained and validated as part of the competition, and perform an additional evaluation using a total of four independently created databases, where resolutions varied from 360p to 2160p, and frame rates from 15--60fps, using realistic coding and bitrate settings.

We found that the model performs well on the independent dataset, with a Pearson correlation of 0.942 and an RMSE of 0.42. We also provide an open-source reference implementation of the described P.1204.3 model, as well as the multi-codec bitstream parser required to extract the input data, which is not part of the standard.



# Are you still watching? Streaming Video Quality and Engagement Assessment in the Crowd

Twelfth International Conference on Quality of Multimedia Experience (QoMEX), May 26 - 28, 2020

**Werner Robitza, Alexander M. Dethof, Steve Göring, Alexander Raake, André Beyer, Tim Polzehl**

We present first results from a large-scale crowdsourcing study in which three major video streaming OTTs were compared across five major national ISPs in Germany. We not only look at streaming performance in terms of loading times and stalling, but also customer behavior (e.g., user engagement) and Quality of Experience based on the ITU-T P.1203 QoE model. We used a browser extension to evaluate the streaming quality and to passively collect anonymous OTT usage information based on explicit user consent. Our data comprises over 400,000 video playbacks from more than 2,000 users, collected throughout the entire year of 2019.

The results show differences in how customers use the video services, how the content is watched, how the network influences video streaming QoE, and how user engagement varies by service. Hence, the crowdsourcing paradigm is a viable approach for third parties to obtain streaming QoE insights from OTTs.

The paper was written together with the TU Ilmenau spin-off AVEQ GmbH, and the Berlin-based company Crowdee GmbH, and it can be downloaded here (<https://aveq.info/resources/>).

## Overview

- A crowdsourcing study on **Over the Top video streaming**
- Data collection throughout 2019
- Desktop PCs and laptops
- Dedicated web browser extension (Chrome & Firefox)
- ~400,000 video playbacks from ~2,000 users

This contribution: An overview of the dataset, first and previously unknown statistics



Population density in Germany (2017), [demografie-portal.de](http://demografie-portal.de)



QoMEX 2020

2

## Prenc - Predict Number Of Video Encoding Passes With Machine Learning

Twelfth International Conference on Quality of Multimedia Experience(QoMEX). Athlone, Ireland. May 2020

**Steve Göring, Rakesh Rao Ramachandra Rao and Alexander Raake**

Video streaming providers spend huge amounts of processing time to get a quality-optimized encoding.

While the quality-related impact may be known to the service provider, the impact on video quality is hard to assess, when no reference is available.

Here, bitstream-based video quality models may be applicable, delivering estimates that include encoding-specific settings. Such models typically use several input parameters, e.g. bitrate, frame-rate, resolution, video codec, QP values and more.

However, for a given bitstream, to determine which encoding parameters were selected, e.g., the number of encoding passes, is not a trivial task.

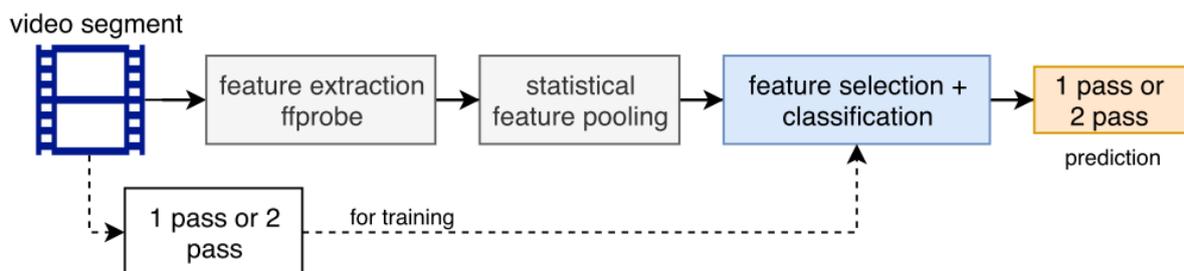
This leads to our following research question: Given an unknown video bitstream, which encoding settings have been used? To tackle this reverse engineering problem, we introduce a system called prenc.

Besides the use in video-quality estimation, such algorithms may also be used in other applications such as video forensics. We prove our concept by applying prenc to distinguish between one- and two-pass encoding.

Starting from modeling the problem as a classification task, estimating bitstream-based features, we further describe a machine learning approach with feature selection to automatically predict the number of encoding passes for a given video bitstream.

Our large-scale evaluation consists of 16 short movie type 4K videos that were segmented and encoded with different settings (resolutions, codecs, bitrates), so that we in total analyzed 131.976 DASH video segments.

We further show that our system is robust, based on a 50\% train and 50\% validation approach without source video overlapping, where we get a classification performance of 65\%~F1 score. Moreover, we also describe the used bitstream-based features in detail, the feature pooling strategy and include other machine learning algorithms in our evaluation.



## Development and evaluation of a test setup to investigate distance differences in immersive virtual environments

2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), May 26 - 28, 2020

**Stephan Fremerey, Muhammad Sami Suleman, Abdul Haq Azeem Paracha and Alexander Raake**

Nowadays, with recent advances in virtual reality technology, it is easily possible to integrate real objects into virtual environments by creating an exact virtual replication and enabling interaction with them by mapping the obtained tracking data of the real to the virtual objects.

The primary goal of our study is to develop a system to investigate distance differences for near-field interaction in immersive virtual environments. In this context, the term distance difference refers to the shift between a real object and the respective replication of the real object in the virtual environment of the same size. This could occur for a number of reasons e.g. due to errors in motion tracking or mistakes in designing the virtual environment. Our virtual environment is developed using the Unity3D game engine, while the immersive contents were displayed on an HTC Vive Pro head-mounted display. The virtual room shown to the user includes a replication of the real testing lab environment, while one of the two real objects is tracked and mirrored to the virtual world using an HTC Vive Tracker.

Both objects are present in the real as well as in the virtual world. To find perceivable distance differences in the near-field, the actual task in the subjective test was to pick up one object and place it into another object.

The position of the static object in the virtual world is shifted by values between 0 and 4 cm, while the position of the real object is kept constant. The system is evaluated by conducting a subjective proof-of-concept test with 18 test subjects.

The distance difference is evaluated by the subjects through estimating perceived confusion on a modified 5-point absolute category rating scale. The study provides quantitative insights into allowable real-world vs. virtual-world mismatch boundaries for near-field interactions, with a threshold value of around 1 cm.

Link to the repository:

[https://github.com/Telecommunication-Telemedia-Assessment/distance\\_differences\\_ives](https://github.com/Telecommunication-Telemedia-Assessment/distance_differences_ives)

## Development and evaluation of a test setup to investigate distance differences in immersive virtual environments

Stephan Fremerey, Muhammad Sami Suleman, Abdul Haq Azeem Paracha and Alexander Raake

Audiovisual Technology Group, TU Ilmenau, Germany

### I. Motivation & Research Questions

- Variety of immersive virtual environments (IVE) developed
- Tracking errors could lead to offset between real and virtual environment → Interaction with objects leading to problems
- Little research has been done on near-field ego-centric distance estimation
- RQ1: Offset value from which the perceived grade of confusion increases?
- RQ2: Influence on task completion time and performance? Learning effect?

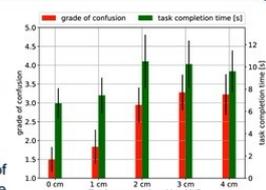
### II. Approach & Test Setup

- Development of test framework using Unity3D
- Conducted subjective test with 18 subjects
- Fulfill simple pick-and-place task at certain offset levels (0 cm, 1 cm, 2 cm, 3 cm, 4 cm)
- Task: "Please place the cylinder inside the other open cylinder as fast as possible."
- Variable position of cylinder on the table, constant position of interactable cylinder
- Between trials: translational directions of the offset levels randomized, order of offsets not
- After task completion: perceived grade of confusion measured using modified version of ACR scale
- Before and after every test: short SSQ (cf. A. Single et al. "Assessing Media QoE, Simulator Sickness and Presence for Omnidirectional Videos with Different Test Protocols")



### III. Results

- In general low short SSQ values (dizziness, headache and eyestrain)
- Every subject fulfilled task at every offset level between real and virtual cylinder
- Lowest amount of confusion (1.5) for condition with no offset
- Starting from 2 cm offset: higher degree of confusion and higher task completion time
- 3 cm and 4 cm offset: grade of confusion not considerably increasing
- 4 cm offset: grade of confusion and task completion time even slightly lower
- Could be related to learning effect of subjects, evolves over time
- Increase of confusion even though order of offsets not randomized



### IV. Conclusions and Future Work

- Publicly available test framework (cf. QR-code)
- IVE can be utilized to measure effects of these distance offsets between physical and virtual objects
- Starting from 2 cm offset: grade of confusion and task completion time increasing
- Results important for researchers and designers creating IVEs
- Offset should be equal to or below 1 cm
- Future work: more subjects, more gradations between offset values, randomize order of offsets, hand and arm tracking

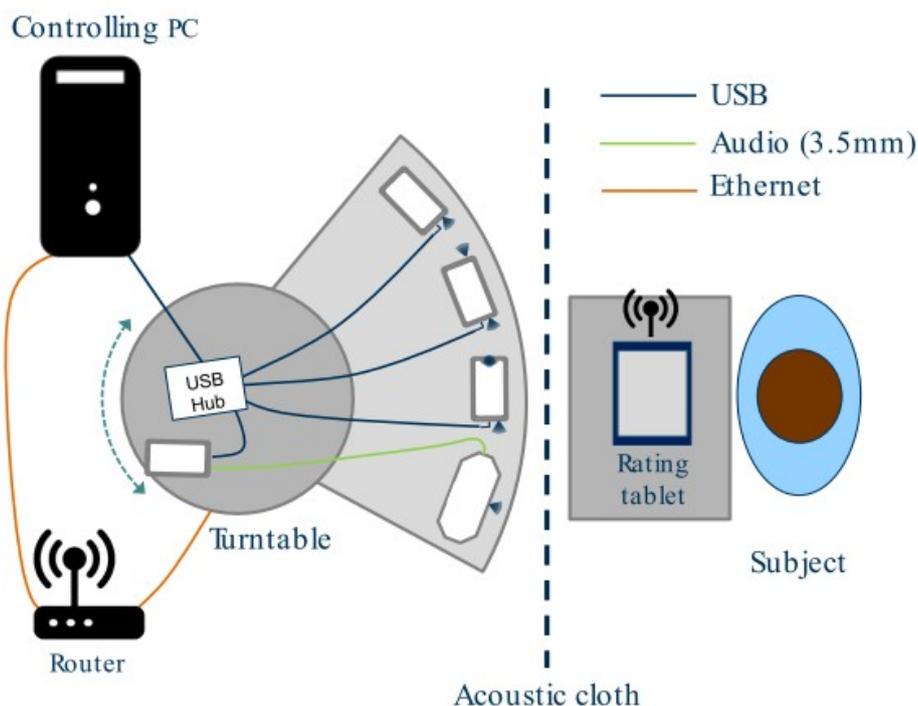


# Let the Music Play: An Automated Test Setup for Blind Subjective Evaluation of Music Playback on Mobile Devices

Twelfth International Conference on Quality of Multimedia Experience (QoMEX), May 2020

**Keller, D.; Raake, A.; Vaalgamaa, M.; Paajanen, E.**

Several methods for subjective evaluation for audio and speech have been standardized over the last years. However, with the advancement of mobile devices such as smartphones and Bluetooth speakers, people listen to music even outside their home environment, when traveling and in social situations. Conventional comparative methodologies are difficult to use for sound-quality evaluation of such devices, since subjects are likely to include other factors such as brand or design. Hence, we propose an automated test setup to evaluate music and audio playback of portable devices with subjects without revealing the devices or interfering with the tests. Furthermore, an identical placement of the devices in front of the listener is crucial to accommodate the individual acoustic directivity of the device. For this purpose, we use a large motorized turntable on which the devices are mounted so that the playback device is automatically moved to the defined position in advance. An enhanced version of rating software avrateNG enables the automatic playout of musical pieces and appropriate turning of the devices to face the listeners. Devices that can automatically be tested using our setup include Android and iOS smartphones, as well as Bluetooth and wired portable speakers. Preliminary user tests were conducted to verify the practical applicability and stability of the proposed setup.

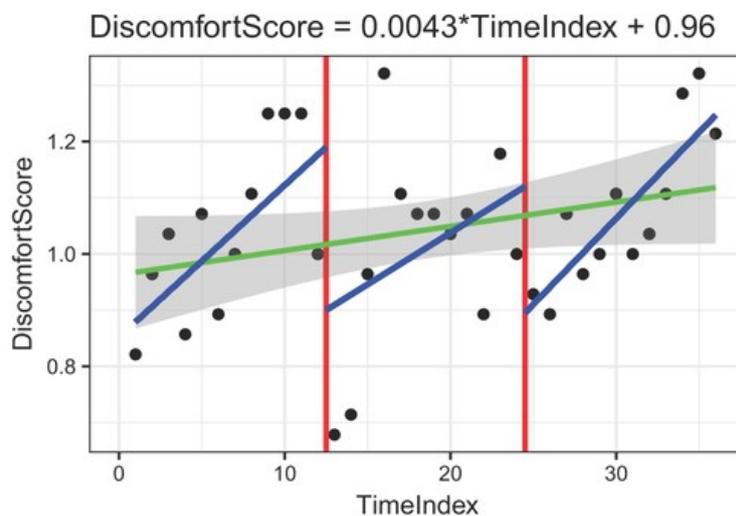


# SiSiMo: Towards Simulator Sickness Modeling for 360° Videos Viewed with an HMD

27th IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), March 2020, Atlanta, USA

**A. Raake, A. Singla, R. R. R. Rao, W. Robitza and F. Hofmeyer**

Users may experience symptoms of simulator sickness while watching 360°/VR videos with Head-Mounted Displays (HMDs). At present, practically no solution exists that can efficiently eradicate the symptoms of simulator sickness from virtual environments. Therefore, in the absence of a solution, it is required to at least quantify the amount of sickness. In this paper, we present initial work on our Simulator Sickness Model SiSiMo including a first component to predict simulator sickness scores over time. Using linear regression of short term scores already shows promising performance for predicting the scores collected from a number of user tests.



## Projekt CYTEMEX

The project is a scientific cooperation between the labs of Audiovisual Technology, Virtual Worlds and Digital Games (Prof. Wolfgang Broll, Faculty of Economics and Media) and Electronic Media Technology (Prof. Karlheinz Brandenburg, Faculty of Electrical Engineering and Information Technology).

The project, funded by the Free State of Thuringia, was co-financed by the European Union within the European Regional Development Fund (ERDF).

[Project Website](#)

## Cyberphysische Telepräsenzsysteme zur Erforschung der Mensch-zu-X-Kommunikation (CYTEMEX)

**Antragsteller-Name:** Technische Universität Ilmenau - Fachgebiet Audiovisuelle Technik

**Beschreibung des Vorhabens:**

Im Vorhaben wird untersucht, welchen Einfluss neuartige Telepräsenztechnologien auf die Kommunikation und soziale Verbundenheit sowie den Erfolg therapeutischer Maßnahmen haben. Cyberphysische VR/AR/MR-basierte Telepräsenzsysteme werden entwickelt und evaluiert.

Das vom Freistaat Thüringen geförderte Vorhaben wurde durch Mittel der Europäischen Union im Rahmen des Europäischen Fonds für regionale Entwicklung (EFRE) kofinanziert.



**EFRE bewegt  
Thüringen**

[www.efre-thueringen.de](http://www.efre-thueringen.de)

Freistaat  
Thüringen



**EFRE**  
EUROPA FÜR THÜRINGEN  
EUROPEAN FUND FOR REGIONAL DEVELOPMENT



EUROPÄISCHE UNION

## ITU-T standard for predicting video quality developed with the significant participation of the AVT department

ITU-T recently consented the P.1204 series of Recommendations titled "Video quality assessment of streaming services over reliable transport for resolutions up to 4K". This work was jointly conducted by Question 14 of Study Group 12 (SG12/Q14) of the ITU-T and the Video Quality Experts Group (VQEG). Overall 9 companies and universities were part of this competition-based development, with the best set of models recommended as standards.

From the official ITU-T SG12 communication it reads:

"The P.1204 Recommendation series describes a set of objective video quality models. These can be used standalone for assessing video quality for 5-10 sec long video sequences, providing a 5-point ACR-type Mean Opinion Score (MOS) output. In addition, they deliver per-1-second MOS-scores that together with audio information and stalling / initial loading data can be used to form a complete model to predict the impact of audio and video media encodings and observed IP network impairments on quality experienced by the end-user in multimedia streaming applications. The addressed streaming techniques comprise progressive download as well as adaptive streaming, for both mobile and fixed network streaming applications."

To date, the P.1204 series of Recommendations comprises four sub-recommendations, namely P.1204 (an introductory document for the whole P.1204 series), P.1204.3 (bitstream-based model with full access to bitstream), P1204.4 (reference-/pixel-based model) and P1204.5 (hybrid bitstream- and pixel-based no-reference) with 2 more sub-recommendations, P1204.1 (meta-data-based) and P1204.2 (meta-data- and video-frame-information-based) planned to be consented by April 2020.

The AVT group of TU Ilmenau in collaboration with Deutsche Telekom were the sole winners in the category which resulted in Recommendation P1204.3 and are co-winners in the category which is planned to result in Recommendations P1204.1 and P1204.2 by April 2020.

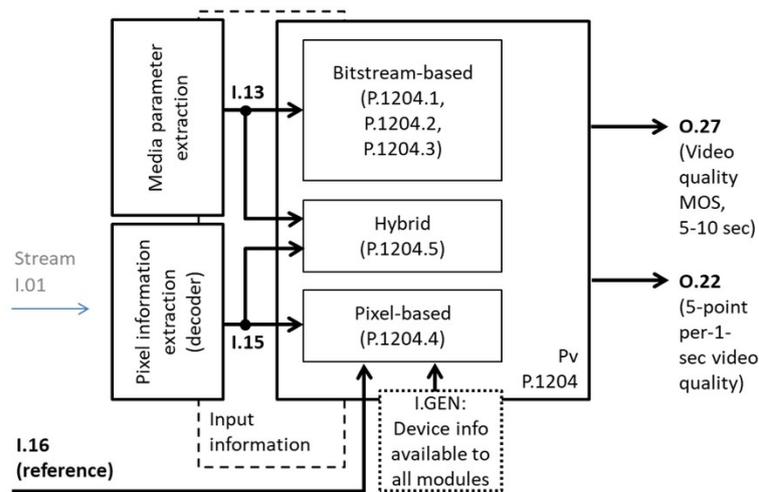
In the official ITU-T SG12 communication it is further stated that:

The consent of the P.1204 model standards marks the first time that video-quality models of all relevant types have been developed and validated within the same standardization campaign. The respective “P.NATS Phase 2” model competition used a total of 13 video-quality test databases for training, and another 13 video-quality test databases for validation. With this comparatively high number of data (more than 5000 video sequences), the resulting standards deliver class-leading video-quality prediction performance.

The published ITU standards:

P.1204: <https://www.itu.int/rec/T-REC-P.1204-202001-P/en>

P.1204.3: <https://www.itu.int/rec/T-REC-P.1204.3-202001-P/en>



## AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1

21st IEEE International Symposium on Multimedia (2019 IEEE ISM), Dec 9 - 11, 2019, San Diego, USA

**Rakesh Rao Ramachandra Rao, Steve Göring, Werner Robitza, Bernhard Feiten, Alexander Raake**

4K television screens or even with higher resolutions are currently available in the market. Moreover video streaming providers are able to stream videos in 4K resolution and beyond. Therefore, it becomes increasingly important to have a proper understanding of video quality especially in case of 4K videos. To this effect, in this paper, we present a study of subjective and objective quality assessment of 4K ultra-high-definition videos of short duration, similar to DASH segment lengths.

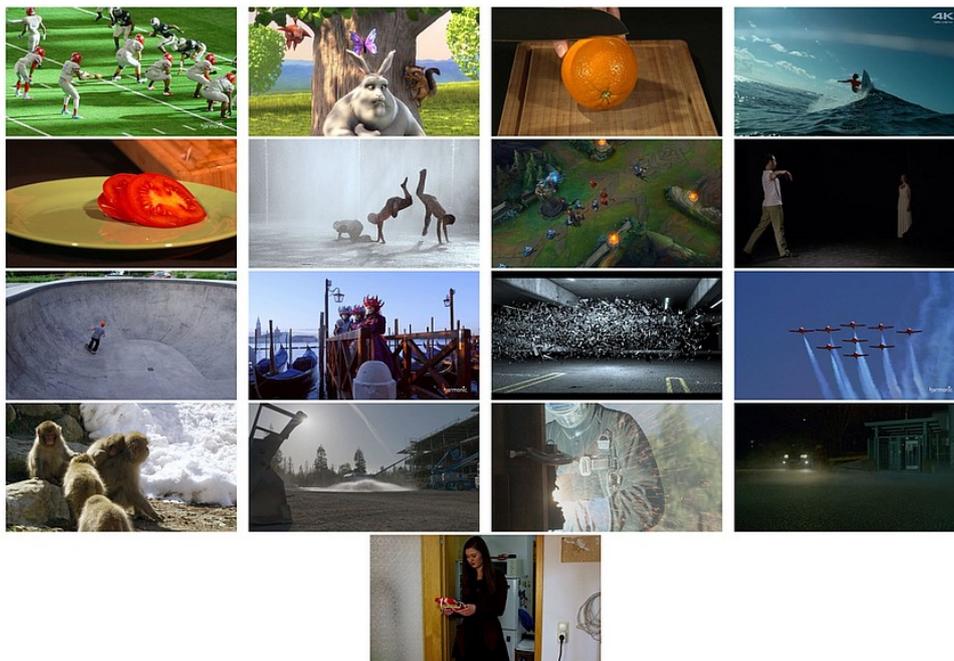
As a first step, we conducted four subjective quality evaluation tests for compressed versions of the 4K videos. The videos were encoded using three different video codecs, namely H.264, HEVC, and VP9. The resolutions of the compressed videos ranged from 360p to 2160p with framerates varying from 15fps to 60fps. All the source 4K contents used were of 60fps. We included low-quality conditions in terms of bitrate, resolution and framerate to ensure that the tests cover a wide range of conditions, and that e.g. possible models trained on this data are more general and applicable to a wider range of real world applications. The results of the subjective quality evaluation

are analyzed to assess the impact of different factors such as bitrate, resolution, framerate, and content.

In the second step, different state-of-the-art objective quality models were applied to all videos and their performance was analyzed in comparison with the subjective ratings, e.g. using Netflix's VMAF. The videos, subjective scores, both MOS and confidence interval per sequence and objective scores are made public for use by the community for further research.

Link to the videos:

<https://github.com/Telecommunication-Telemedia-Assessment/AVT-VQDB-UHD-1>



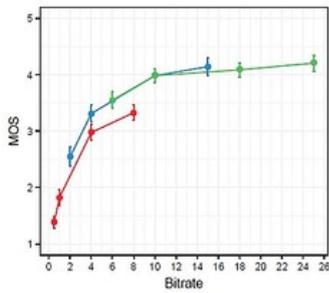
## Comparison of Subjective Quality Test Methods for Omnidirectional Video Quality Evaluation

21st IEEE International Workshop on Multimedia Signal Processing (MMSP), September 2019, Kuala Lumpur, Malaysia

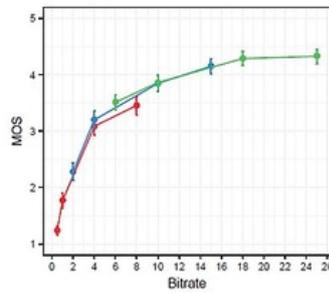
**A. Singla, W. Robitza and A. Raake**

The test methods recommended by the International Telecommunication Union (ITU) for assessing 2D video quality are often used for evaluating omnidirectional / 360° videos. In this paper, we compare the performance of three different test methods, Absolute Category Rating (ACR), a modified version of ACR (M-ACR) with double presentation of the test stimulus, and DSIS (Double Stimulus Impairment Scale), based on the statistical reliability, assessment time and simulator sickness. Different settings were used for HEVC encoding of five 360° source videos of 10 s duration. Results indicate that DSIS is statistically more reliable with higher resolving power, followed by M-ACR and ACR. We found that simulator sickness increases with time, but can be reduced by taking breaks in between the test sessions. The results for simulator sickness are compared across test methods and with similar tests conducted under different contextual conditions. We also recorded and analyzed the exploration behaviour of the users. Apart from the methodological findings, the test re-

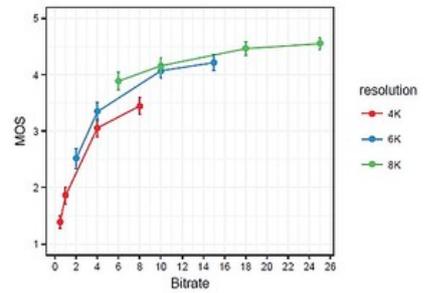
sults provide insights into video quality for different resolution and encoding settings (“bitrate ladders”). These may be useful for choosing appropriate representations in the context of HTTP-based adaptive streaming in case of full-frame streaming.



(a) ACR



(b) M-ACR



(c) DSIS

## Best Paper Award

Dominik Keller (AVT Group), Tamara Seybold (ARRI Munich), Janto Skowronek (former AVT Group) and Alexander Raake (AVT Group) got the Best Paper Award at the 11th International Conference on Quality of Multimedia Experience (QoMEX 2019) in Berlin.

You find the abstract of the article below.



## Prizes for Graduates of the AVT Group

For the second time, the Förderverein Elektrotechnik und Informationstechnik e. V. Ilmenau (Association for the Promotion of Electrical Engineering and Information Technology Ilmenau) in conjunction with the Department of Electrical Engineering and Information Technology of the TU Ilmenau presented its award for outstanding theses. The three endowed prizes honor the achievements of the students during the exmatriculation ceremony at the end of June. Fortunately, two master theses of the AVT group which were carried out with industrial partners were honored and

awarded as outstanding due to their high degree of interdisciplinarity and scientific character as well as their execution.

We congratulate the award winners Anton Schubert, who has worked on the implementation of a compressed broadband audio codec for driver communication in motor sports, and Dominik Keller, who has worked on identification and analysis of texture dimensions in motion pictures using sensory evaluation techniques.



The winners Dominik Keller and Anton Schubert with the chairman of the Förderverein Prof. Seitz.

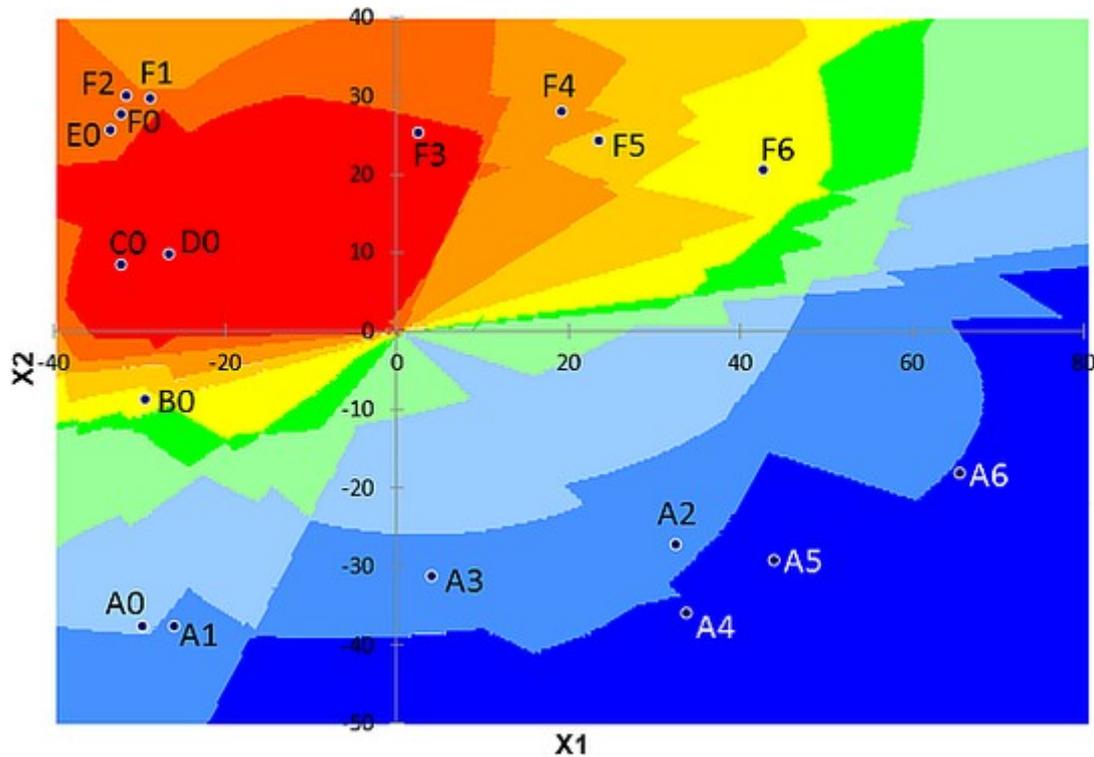
## Assessing Texture Dimensions and Video Quality in Motion Pictures using Sensory Evaluation Techniques

**Dominik Keller, Tamara Seybold, Janto Skowronek und Alexander Raake**

The paper resulting from the cooperation of members of the Audiovisual Technology Group and Scientific and Engineering Academy Award winner ARRI (Arnold & Richter Cine Technik) received Best Paper Award at this year's 11th Int. Conference on Quality of Multimedia Experience (QoMEX 2019).

The quality of images and videos is usually examined with well-established subjective tests or instrumental models. These often target content transmitted over the internet, such as streaming or videoconferences and address the human preferential experience. In the area of high-quality motion pictures, however, other factors are relevant. These mostly are not error-related but aimed at the creative image design, which has gained comparatively little attention in image and video quality research. To determine the perceptual dimensions underlying movie-type video quality, we combine sensory evaluation techniques extensively used in food assessment – Degree of Difference test and Free Choice Profiling – with more classical video quality tests. The main goal of this research is to analyze the suitability of sensory evaluation methods for high-quality video assessment. To understand which features in motion pictures are recognizable and critical to quality, we address the example of image texture properties, measuring human perception and preferences with a panel of image-quality experts. To this aim, different capture settings were simulated applying sharpening filters as well as digital and analog noise to exemplary source sequences. The evalu-

ation, involving Multidimensional Scaling, Generalized Procrustes Analysis as well as Internal and External Preference Mapping, identified two separate perceptual dimensions. We conclude that Free Choice Profiling connected with a quality test offers the highest level of insight relative to the needed effort. The combination enables a quantitative quality measurement including an analysis of the underlying perceptual reasons.



External Preference Mapping results: Best ratings for stimuli of low noise and medium-high sharpness (Landscape scene)

# Impact of Various Motion Interpolation Algorithms on 360° Video QoE

Stephan Fremerey, Frank Hofmeyer, Steve Göring and Alexander Raake

Audiovisual Technology Group, TU Ilmenau, Germany

## I. Motivation

- 360° video QoE dependent on many technical and perceptual factors
- Refresh rate of HMDs mostly at 90Hz
- 360° videos mostly available at 25/30fps → mismatch leading to significant quality issues
- 90fps content playback smooth, other versions suffer from confirmed issues (black frame insertion, flickering, stuttering)

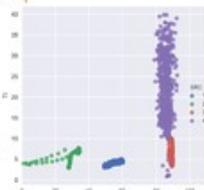


## II. Research Questions and Key Aspects

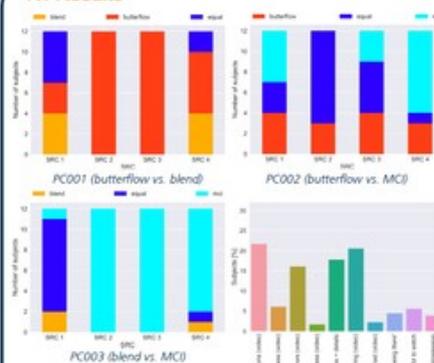
- Result from previous study: Motion Interpolation (MI) is improving QoE of 360° videos (cf. Hofmeyer, Fremerey et al.: "Impacts of internal HMD Playback Processing on Subjective Quality Perception", presented at HVEI 2019)
- Key aspect: Which MI algorithm to choose for which 360° video?

## III. Approach & Test Setup

- 12 video experts, four 360° SRCs, 20s
- SRC selection based on SI/TI
  - Source: 30fps
  - Interpolation: 90fps
  - Res.: 3840x1920
- 3 MI algos:
  - ffmpeg Motion Compensated Interpolation (MCI)
  - ffmpeg blend
  - butterflow
- Test method: paired comparison (video 1 vs. 2)
  - PC001: butterflow vs. blend
  - PC002: butterflow vs. MCI
  - PC003: blend vs. MCI
- HMD: HTC Vive Pro
- Encoding: ffmpeg libx265, CRF=0 (visual lossless)
- Pre-defined questionnaire for feature analysis



## IV. Results



### Results Pair Comparison Test

	PC001	PC002	PC003
SRC 1	Equal	Equal	Equal
SRC 2	Butterflow	Butterflow (slightly)	MCI
SRC 3	Butterflow	Equal	MCI
SRC 4	Butterflow (slightly)	MCI	MCI

- Feature analysis: General video distortions (≈22%), flickering (≈21%) and perception of faces/important details (≈17%) are essential features for quality preference rating of subjects

## V. Conclusions and Future Work

- Slow + medium movements: butterflow providing slightly higher QoE than MCI
- Fast + sudden movements: ffmpeg MCI is best
- MCI has less mosquito artefacts than butterflow
- blend very fast, but not ideal for 360° video
- Future work: Test with more (and naive) subjects, more contents, recommend certain MI algorithms

In a study presented at the QoMEX 2019 conference, we compare the impact of various motion interpolation (MI) algorithms on 360° video Quality of Experience (QoE). For doing so, we conducted a subjective test with 12 video expert viewers, while a pair comparison test method was used. We interpolated four different 20 s long 30 fps 360° source contents to the native 90 Hz refresh rate of popular Head-Mounted Displays using three different MI algorithms. Subsequently, we compared these 90 fps videos against each other to investigate the influence on the QoE. Regarding the algorithms, we found out that ffmpeg blend does not lead to a significant improvement of QoE, while MCI and butterflow do so. Additionally, we concluded that for 360° videos containing fast and sudden movements, MCI should be preferred over butterflow, while butterflow is more suitable for slow and medium motion videos. While comparing the time needed for rendering the 90 fps interpolated videos, ffmpeg blend is the fastest, while MCI and butterflow need much more time.

## Assessing Media QoE, Simulator Sickness and Presence for Omnidirectional Videos with Different Test Protocols

**A. Singla, R. R. R. Rao, S. Göring and A. Raake**

Published in 26th IEEE Conference on Virtual Reality and 3D User Interfaces, March 2019, Osaka, Japan

QoE for omnidirectional videos comprises additional components such as simulator sickness and presence. In this paper, a series of tests is presented comparing different test protocols to assess integral quality, simulator sickness and presence for omnidirectional videos in one test run, using the HTC Vive Pro as head-mounted display. For quality ratings, the five-point ACR scale was used. In addition, the well-established Simulator Sickness Questionnaire and Presence Questionnaire methods were used, once in a full version, and once with only one single integral scale, to analyze how well presence and simulator sickness can be captured using only a single scale.



Ashutosh Singla while presenting his poster at the IEEE VR conference in Japan

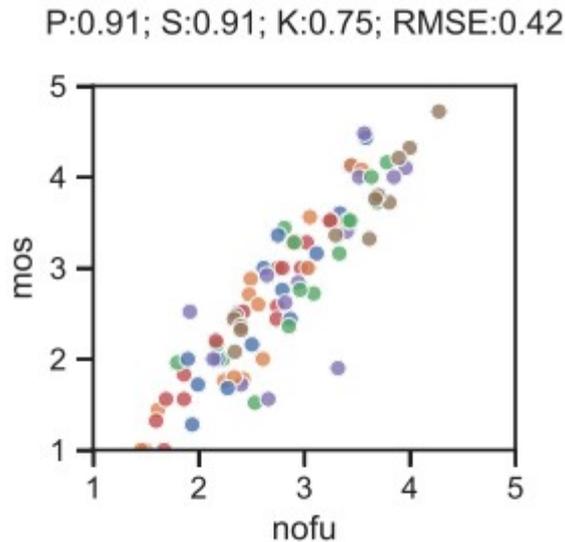
## nofu - A Lightweight No-Reference Pixel Based Video Quality Model for Gaming Content

**Steve Göring, Rakesh Rao Ramachandra Rao, Alexander Raake**

Eleventh International Conference on Quality of Multimedia Experience (QoMEX) (QoMEX 2019). Berlin, Germany. June 2019

Popularity of streaming services for gaming videos has increased tremendously over the last years, e.g. Twitch and Youtube Gaming. Compared to classical video streaming applications, gaming videos have additional requirements. For example, it is important that videos are streamed live with only a small delay. In addition, users expect low stalling, waiting time and in general high video quality during streaming, e.g. using http-based adaptive streaming. These requirements lead to different challenges for quality prediction in case of streamed gaming videos. We describe newly

developed features and a no-reference video quality machine learning model, that uses only the recorded video to predict video quality scores. In different evaluation experiments we compare our proposed model nofu with state-of-the-art reduced or full reference models and metrics. In addition, we trained a no-reference baseline model using brisque+niqe features. We show that our model has a similar or better performance than other models. Furthermore, nofu outperforms VMAF for subjective gaming QoE prediction, even though nofu does not require any reference video.



scatter\_plot\_mos\_nofu: results for gaming dataset and subjective score prediction

## deimeq – A Deep Neural Network Based Hybrid No-reference Image Quality Model

7th European Workshop on Visual Information Processing (EUVIP), Tampere (Finland), 26 - 28 November 2018 (<http://www.tut.fi/euvip2018/>)

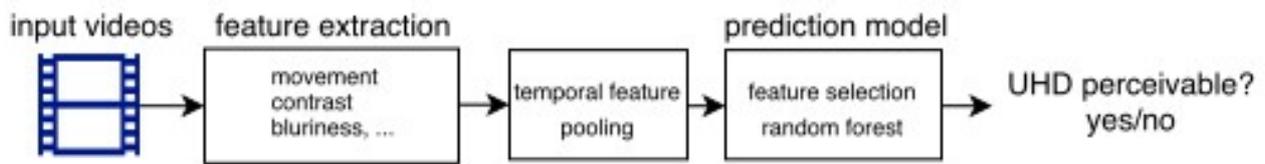
**Steve Göring, Alexander Raake**

Current no reference image quality assessment models are mostly based on hand-crafted features (signal, computer vision, . . . ) or deep neural networks. Using DNNs for image quality prediction leads to several problems, e.g. the input size is restricted; higher resolutions will increase processing time and memory consumption. Large inputs are handled by image patching and aggregation a quality score. In a pure patching approach connections between the sub-images are getting lost.

Also, a huge dataset is required for training a DNN from scratch, though only small datasets with annotations are available. We provide a hybrid solution (deimeq) to predict image quality using

DNN feature extraction combined with random forest models. Firstly, deimeq uses a pre-trained DNN for feature extraction in a hierarchical sub-image approach, this avoids a huge training dataset. Further, our proposed sub-image approach circumvents a pure patching, because of hierarchical connections between the sub-images. Secondly, deimeq can be extended using signal-based features from state-of-the art models. To evaluate our approach, we choose a strict cross-dataset

evaluation with the Live-2 and TID2013 datasets with several pre-trained DNNs. Finally, we show that deimeq and variants of it perform better or similar than other methods.



## Analyze And Predict the Perceptibility of UHD Video Contents

Human Vision and Electronic Imaging 2019, Burlingame (California USA), 13 - 17 January 2019 ([http://www.imaging.org/site/IST/IST/Conferences/EI/Symposium\\_Overview.aspx](http://www.imaging.org/site/IST/IST/Conferences/EI/Symposium_Overview.aspx))

**Steve Göring, Julian Zebelein, Simon Wedel, Dominik Keller, Alexander Raake**

720p, Full-HD, 4K, 8K, ..., display resolutions are increasing heavily over the past time. However many video streaming providers are currently streaming videos with a maximum of 4K/UHD-1 resolution. Considering that normal video viewers are enjoying their videos in typical living rooms, where viewing distances are quite large, the question arises if more resolution is even recognizable. In the following paper we will analyze the problem of UHD perceptibility in comparison with lower resolutions. As a first step, we conducted a subjective video test, that focuses on short uncompressed video sequences and compares two different testing methods for pairwise discrimination of two representations of the same source video in different resolutions.

We selected an extended stripe method and a temporal switching method. We found that the temporal switching is more suitable to recognize UHD video content. Furthermore, we developed features, that can be used in a machine learning system to predict whether there is a benefit in showing a given video in UHD or not.

Evaluating different models based on these features for predicting perceivable differences shows good performance on the available test data. Our implemented system can be used to verify UHD source video material or to optimize streaming applications.

## Prize for P. Lebreton, S. Fremerey and A. Raake

During the "Grand Challenge on Salient 360!" at this year's IEEE International Conference on Multimedia and Expo (ICME) in San Diego, Dr. Pierre Lebreton (formerly TU Ilmenau, now Zhejiang University, China), Stephan Fremerey and Prof. Dr. Alexander Raake received the prize for second place in the category "Prediction of Head Saliency for Images" and fourth place in the category "Prediction of Head Saliency for Videos". (<http://www.icme2018.org>)



## New work on VR, image appeal, and video streaming presented at IEEE QoMEX, ACM MMSys

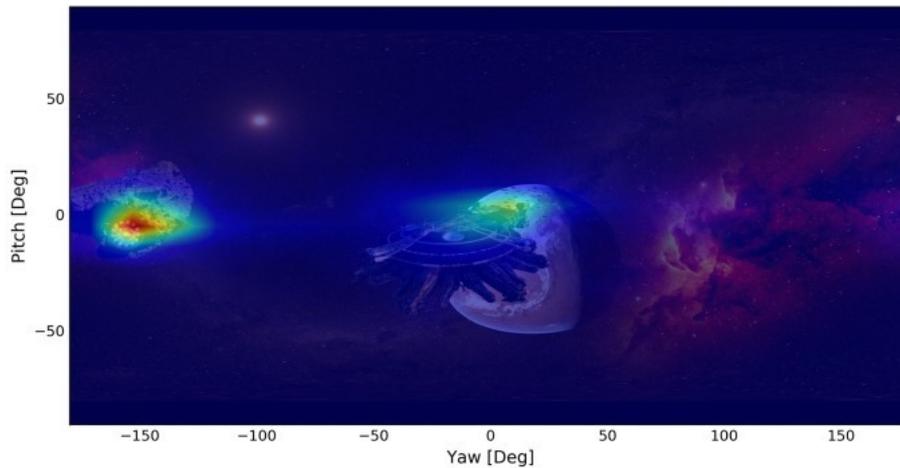
At this year's ACM MMSys conference in Amsterdam, Stephan Fremerey has published a study and a related Open Source dataset and software. The study was done in collaboration with ARTE G.E.I.E. while the research goal was to get insights into the exploration behavior of 48 participants watching 20 different 30s long 360° videos in a task-free scenario. The dataset containing the Simulator Sickness Questionnaire scores, the head rotation data and the software to record and evaluate the respective data are published as Open Source (see <https://github.com/Telecommunication-Telemedia-Assessment/AVTrack360>).

Steve Göring's work on image appeal was presented at IEEE QoMEX. The core idea of the paper is to train a model for image liking prediction based on a crawled dataset of 80k images from a photo sharing platform. The used features are based on the image, social network analysis, comment analysis and other provided meta-data of images from such platforms.

On the topic of video streaming, Werner Robitza has published a study at QoMEX. The research was conducted in collaboration with Deutsche Telekom, where the quality of YouTube streams was measured under different bandwidth conditions. The paper shows the influence of different measurement scenarios on the measured key performance indicators and quality according to ITU-T P.1203, a standard for HTTP Adaptive Streaming quality estimation.

Together with academic and industry partners from Ericsson, Deutsche Telekom, TU Berlin, NTT Corporation, and NETSCOUT Systems, an open dataset and software for the ITU-T P.1203 standard

was published at the ACM MMSys conference. The software can be used freely for non-commercial research purposes. You can find out more about the group's work on P.1203 [on this page](#).



Example for a heatmap showing the head rotation data captured for all participants for one sequence

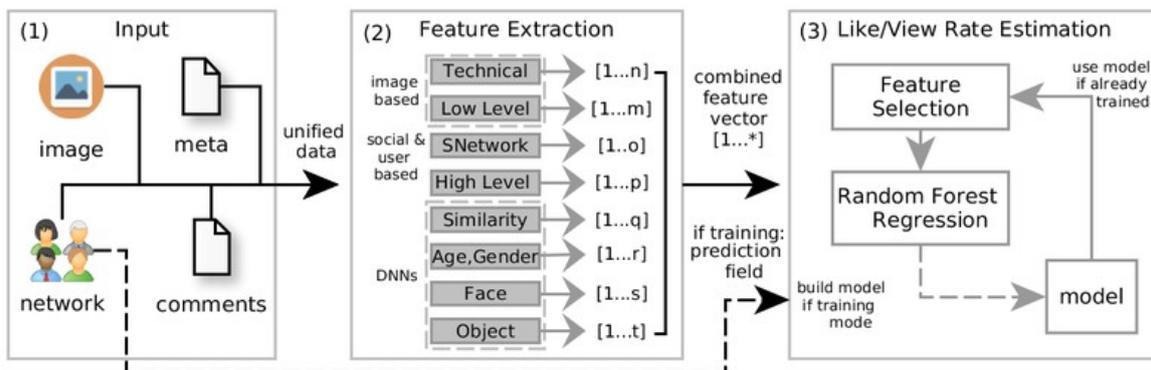
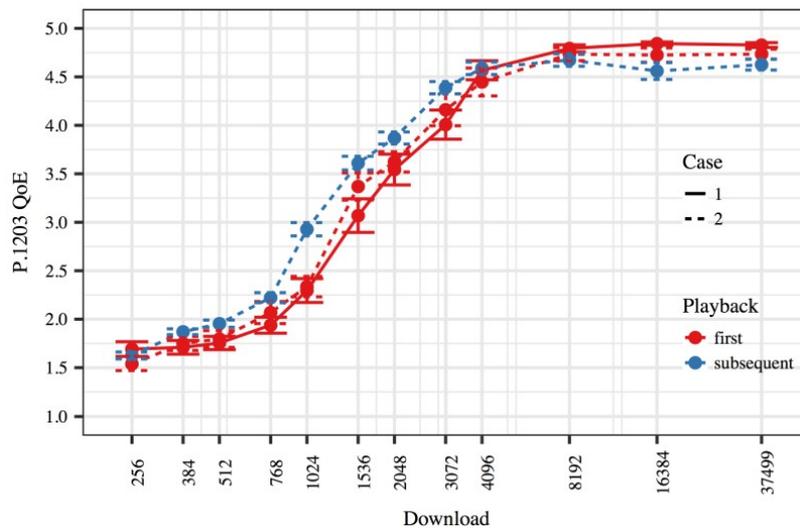


Image Appeal Prediction Pipeline



YouTube quality according to ITU-T P.1203 dependent on download speed and other factors.

## IMT at the summer party of the Thuringian State Representation in Berlin

As in the previous year, the department was represented with demonstrations of its current research topics at the summer party of the Thuringian State Representation in Berlin. At the event in the afternoon and evening of June 26, 2018, interested visitors were able to find out about techniques of virtual reality and the mechanisms used for quality estimation at the booth of the TU Ilmenau. Also, after watching 360° videos, they could see which parts of the video they had actually explored.

Prominent visitor and interested guest was Thuringian Minister-President Bodo Ramelow. Prof. Raake explained the research goals in the field of VR technology.

Prof. Raake was supported during the event by Stephan Fremerey, Matthias Döring and Paul Krabbes from the Institute for Media Technology.

[Video \(YouTube\)](#)



Prof. Dr. Alexander Raake in conversation with the Thuringian Prime Minister Bodo Ramelow (Photo: M. Döring)

## FKTG Honorary Membership for Prof. Schade

Prof. Dr.-Ing. Hans-Peter Schade, who was the head of the AVT laboratory from 2002 to 2015, was appointed honorary member of the Fernseh- und Kinotechnischen Gesellschaft (FKTG) - Television and Cinematographic Society - at the 28th FKTG Conference on June 4-6, 2018 in Nuremberg.



Prof. Schade (center) with the managing director Jürgen Burghardt and the chairman Siegfried Föbel of the FK TG

## STEEM project was finished

In May 2018, the STEEM project (Speech Transmission End-to-End Monitoring) was successfully completed. Based on a large number of conversational tests, in the lab has developed an improved model for predicting the perceived quality of IP telephony. In addition to existing models, the influence of background noise and terminal device characteristics were included in the model. This model was handed over to the project partner HEAD acoustics. The solution will find its way into the products of the company in order to better enable network operators and component manufacturers to predict the voice quality perceived by the customer and to optimize their products and services with regard to customer requirements.

Partial results of the conversational tests carried out in Ilmenau were presented at the DAGA 2018. From this publication also comes the following figure.



Conversation test in the AVT test lab with different terminals and a system for defined background noise.

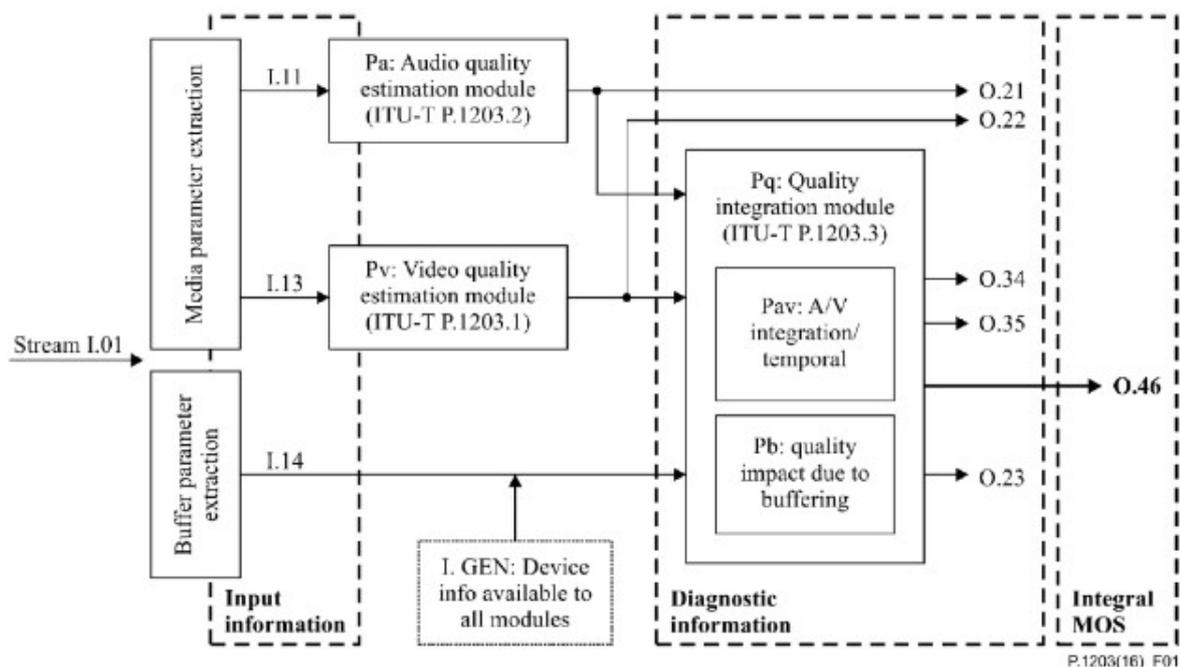
## Award for P. Lebreton and A. Raake at ICME 2017

Dr. Pierre Lebreton (former TU Ilmenau, now Zhejiang University, China) and Prof. Dr. Alexander Raake got the "Award for Best Head Movement Prediction" at the IEEE International Conference on Multimedia and Expo (ICME) 2017 in Hong Kong.

<http://www.icme2017.org/>



## P.NATS Phase 1 - TU Ilmenau part of all winning groups



Building blocks of the ITU-T P.1203 model

The International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) has recently published the P.1203 series of recommendations. This standard is the first which allows to measure the quality of different streaming services such as YouTube, Netflix etc. The standardized models were developed in a competition within the ITU-T Study Group 12 (SG12) which was previously referred to as the P.NATS (Parametric Non-intrusive Assessment of TCP-based multimedia Streaming quality) consisting of seven participating companies. The standardized models are available in four modes of operation depending on the different levels of media-related information used for analysis which in turn reflect the different types of media stream encryption. The models are developed to assess the quality of streaming services up to HD resolution. A modular model framework has been used to develop the models. This modular architecture is as shown in the figure.

**TU Ilmenau in collaboration with Deutsche Telekom has been part of the winning groups for all the four modes.**

Sources:

Raake, A., Garcia, M.N., Robitza, W., List, P., Göring, S. and Feiten, B., 2017, May.

A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P. 1203.1. In *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference* on (pp. 1-6). IEEE. <http://ieeexplore.ieee.org/abstract/document/7965631/>

Robitza, W., Garcia, M.N. and Raake, A., 2017, May.

A modular HTTP adaptive streaming QoE model—Candidate for ITU-T P. 1203 (“P. NATS”). In *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference* on (pp. 1-6). IEEE. <http://ieeexplore.ieee.org/abstract/document/7965689/>

Links to the standardization documents:

- 1) [Recommendation P.1203] - <https://www.itu.int/rec/T-REC-P.1203-201611-I/en>
- 2) [P.1203.1] - <https://www.itu.int/rec/T-REC-P.1203.1/en>
- 3) [P.1203.2] - <https://www.itu.int/rec/T-REC-P.1203.2/en>
- 4) [P.1203.3] - <https://www.itu.int/rec/T-REC-P.1203.3/en>

Other links:

- 1) <http://newslog.itu.int/archives/1477>
- 2) <http://www.laboratories.telekom.com/public/English/Newsroom/news/Pages/Neuer-ITU-Standard.aspx>

## QoMEX 2017



### QoMEX 2017

Erfurt, May 31 – June 02, 2017

In May 2017, the Audiovisual Technology Group of TU Ilmenau organized the 9th International Conference QoMEX 2017 in Erfurt. The conference brought together leading experts from academia and industry to present and discuss current research on multimedia quality, Quality of Experience (QoE) and user experience.

More information is available on the website: [www.gomex2017.org](http://www.gomex2017.org)

## Book "Quality of Experience"

The book "Quality of Experience" appeared in 2014. Alexander Raake is the co-editor, and many other former members of the AIPA team are co-authors of different chapters. The book:

- develops the definition of "Quality of Experience" for telecommunication services and applies it to various fields of communication and media technology
- provides examples and guidelines for many fields of application
- • was written by well-known experts in the field

[Information at Springer-Verlag](#)

