TECHNISCHE UNIVERSITÄT
ILMENAU

Quality Taxonomy for Scalable Algorithms of Free Viewpoint Video Objects

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät für Elektrotechnik und Informationstechnik
Technische Universität Ilmenau

von Dipl.-Ing.(FH) Sara Kepplinger

Gutachter Prof. Dr.-Ing. Karlheinz Brandenburg
Prof. Dr. Ingrid Heynderickx
Prof. Dr. Patrick Le Callet

Tag der Einreichung: 11. Januar 2017
Tag der wissenschaftlichen Aussprache: 25. Oktober 2017

urn:nbn:de:gbv:ilm1-2017000558

Declaration

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die vorliegende Arbeit wurde am Fachgebiet Elektronische Medientechnik des Instituts für Medientechnik an der Technischen Universität Ilmenau unter wissenschaftlicher Betreuung angefertigt. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise geholfen:

1. Unentgeltlich: Herr Dr. Christian Weigel (Entwicklung der technischen Realisierungsprozesse und Erstellung der in Kapitel 2 vorgestellten und in der Arbeit untersuchten Teststimuli im Rahmen des DFG-Forschungsprojektes *Skalierbare Algorithmen für 3D Videoobjekte unter Berücksichtigung subjektiver Qualitätsfaktoren.*)
2. Entgeltlich: Franziska Reif und Jana Kaiser (Korrekturlesen Englische Sprache)

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt. Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch angesehen wird und gemäß §7 Abs.10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Ilmenau, December 21, 2017 Sara Kepplinger

Abstract

The thesis intends to make a contribution to the quality assessment of free viewpoint video objects within the context of video communication systems. The current work analyzes opportunities and obstacles, focusing on users' subjective quality of experience in this special case. Quality estimation of emerging free viewpoint video object technology in video communication has not yet been assessed and adequate approaches are missing. The challenges are to define factors that influence quality, to formulate an adequate measure of quality, and to link the quality of experience to the technical realization within an undefined and ever-changing technical realization process. There are two advantages of interlinking the quality of experience with the quality of service: First, it can benefit the technical realization process, in order to allow adaptability (e.g., based on systems used by the end users). Second, it provides an opportunity to support scalability in a user-centered way, e.g., based on a cost or resources limitation. The thesis outlines the theoretical background and introduces a user-centered quality taxonomy in the form of an interlinking model. A description of the related project *Skalalgo3d* is included, which offered a framework for application. The outlined results consist of a systematic definition of factors that influence quality, including a research framework, and evaluation activities involving more than 350 participants. The thesis includes the presentation of quality features, defined by evaluations of free viewpoint video object quality, for video communication application. Based on these quality features, a model that links these results with the technical creation process, including a formalized quality measure, is presented. Based on this, a flow chart and slope field are proposed. These intend the visualization of these potential relationships and may work as a starting point for further investigations thereon and to differentiate relations in form of functions.

Kurzfassung

Diese Dissertation beabsichtigt einen Beitrag zur Qualitätsbeurteilung von Algorithmen für Bildanalyse und Bildsynthese im Anwendungskontext Videokommunikationssysteme zu leisten. In der vorliegenden Arbeit werden Möglichkeiten und Hindernisse der nutzerzentrierten Definition von subjektiver Qualitätswahrnehmung in diesem speziellen Anwendungsfall untersucht. Qualitätsbeurteilung von aufkommender Visualisierungs-Technologie und neuen Verfahren zur Erzeugung einer dreidimensionalen Repräsentation unter der Nutzung von Bildinformation zweier Kameras für Videokommunikationssysteme wurde bisher noch nicht umfangreich behandelt und passende Ansätze dazu fehlen. Die Herausforderungen sind es qualitätsbeeinflussende Faktoren zu definieren, passende Maße zu formulieren, sowie die Qualitätsevaluierung mit den Erstellungsalgorithmen, welche noch in Entwicklung sind, zu verbinden. Der Vorteil der Verlinkung von Qualitätswahrnehmung und Servicequalität ist die Unterstützung der technischen Realisierungsprozesse hinsichtlich ihrer Anpassungsfähigkeit (z.B. an das vom Nutzer verwendete System) und Skalierbarkeit (z.B. Beachtung eines Aufwands- oder Ressourcenlimits) unter Berücksichtigung des Endnutzers und dessen Qualitätsanforderungen. Die vorliegende Arbeit beschreibt den theoretischen Hintergrund und einen Vorschlag für eine Qualitätstaxonomie als verlinkendes Modell. Diese Arbeit beinhaltet eine Beschreibung des Projektes *Skalalgo3d*, welches den Rahmen der Anwendung darstellt. Präsentierte Ergebnisse bestehen aus einer systematischen Definition von qualitätsbeeinflussenden Faktoren inklusive eines Forschungsrahmens und Evaluierungsaktivitäten die mehr als 350 Testteilnehmer inkludieren, sowie daraus heraus definierte Qualitätsmerkmale der evaluierten Qualität der visuellen Repräsentation für Videokommunikationsanwendungen. Ein darauf basierendes Modell um diese Ergebnisse mit den technischen Erstellungsschritten zu verlinken wird zum Schluss anhand eines formalisierten Qualitätsmaßes präsentiert. Ein Flussdiagramm und ein Richtungsfeld zur grafischen Annäherung an eine differenzierbare Funktion möglicher Zusammenhänge werden daraufhin für weitere Untersuchungen vorgeschlagen.

Acknowledgement

Ich freue mich darüber, als Absolventin der Fachhochschule Salzburg an der Technischen Universität Ilmenau die Möglichkeit bekommen zu haben eine Dissertation zu schreiben. In diesem Zusammenhang bedanke ich mich bei Herrn Prof. Groß und Herrn Prof. Brandenburg als Eignungsprüfer zur Annahme als Doktorandin und Herrn Prof. Brandenburg im Besonderen als Doktorvater. Ich möchte mich bei meinem Kollegium des Fachgebietes Elektronische Medientechnik am Institut für Medientechnik und am Fraunhofer-Institut für Digitale Medientechnologie IDMT für die Unterstützung während des Berufsall"-tags bedanken. Dominik Strohmeier danke ich für die Bereitschaft zu diskutieren und für fachmännisches Feedback, ebenso Judith Liebetrau und Thomas Sporer. Johannes Nowak und Christian Weigel danke ich für die LaTeX Tipps und für's Lesen. Meinem Kollegen Christian Weigel danke ich insbesondere für die Zusammenarbeit im Rahmen des Projektes Skalalgo3d. Marianna Obrist, Regina Bernhaupt und meinen ehemaligen Kollegen am ICT&S Center danke ich dafür, dass sie das Forschungsinteresse in mir erblühen lassen haben. David Wilfinger und Christian Valuch danke ich für ihr Feedback. Joachim und Elisabeth Ardelt danke ich für ihre liebevolle Unterstützung während der Anfangsphase der Arbeit. Bei der Universität Salzburg bedanke ich mich für die Möglichkeit eine Studie mit über 300 Psychologiestudierenden durchführen zu können. Dem Doctoral Consortium inklusive Teilnehmer, Reviewer und der Community der EuroITV 2011 danke ich ganz besonders für ihr Feedback, ihre motivierenden Worte, und das Interesse an meiner Arbeit, wonach der Stein so richtig ins Rollen gekommen ist. Ingrid Heynderickx und Patrick Le Callet danke ich für die Bereitschaft Gutachter zu sein und für ihr wertvolles Feedback zum Gelingen der Arbeit: I thank you both very much for your support. Markus Mehnert danke ich für die immerwährende Diskussionsbereitschaft und die mentale und sportliche Unterstützung zu Arbeitsbeginn. Stephan, Dietmar, Mama, Oma, Gerda und Hubert mit Familie, Conny, und Steffi danke ich für ihr offenes Ohr, ihren Beistand und vor allem auch für ihr Interesse. Für seinen Einsatz als Lektor danke ich vor allem meinem Bruder David.

Contents

1 Introduction

During the usage of a video communication system the user usually watches the visual representation of the conversation partner on the display. The quality of this visual representation may vary. Such variations may depend on the kind of data transmission, on the coding, on the way the representation has been created (e.g., camera equipment used, on the algorithms used for image processing), on the available hard- and software resources, and so on. A further issue may be the lack of eye contact during the video communication. This becomes an issue because the user looks at the display and not into the camera (and the user's conversation partner does the same). The investigated application is a concept to support eye contact in video communication by the usage of three-dimensional video objects (3DVO). Chapter 2 of this thesis provides a description of 3DVO that have been created to support eye contact in video communication. These were realized within the research project *Scalable algorithms for 3D video objects under consideration of subjective quality factors* (Skalalgo3d) which aimed to solve the problem of an existing misalignment between the camera and the communication window. This problem was investigated with the help of a virtual view of a camera placed at the position of this communication window. This virtual view is modeled as an algorithm chain of processing steps (see Weigel and Treutner [1]). The presented approach uses a Pixel (px)-based representation of the extracted Three-Dimensional (3D) information, namely a disparity map extracted from a stereo camera setup. The method consists of a chain of processing steps which are described in Section 2.3.2. This chain of processing steps is investigated by quality assessment and defines the technical factors, the quality elements (see also 3.2.1), in this thesis.

Figure 1.1 shows visual representations of a possible video communication use case (without background). The left part (a.) shows a well known scenario, a representation of the conversation partner recorded with an ordinary web camera and without any further intended manipulation (i.e., no horizontal correction of the object, no pre-processing or color correction). The middle section (b.) shows a similar representation but with the conversation partner looking directly into the web camera, leading to the impression that eye contact exists. The right part of the figure (c.) shows a manipulated representation (i.e., a 3DVO). This has been created with the aim of supporting eye contact but includes many errors that influence quality perception. Here, the viewer sees rather extraordinary and unfamiliar distortions and mistakes.

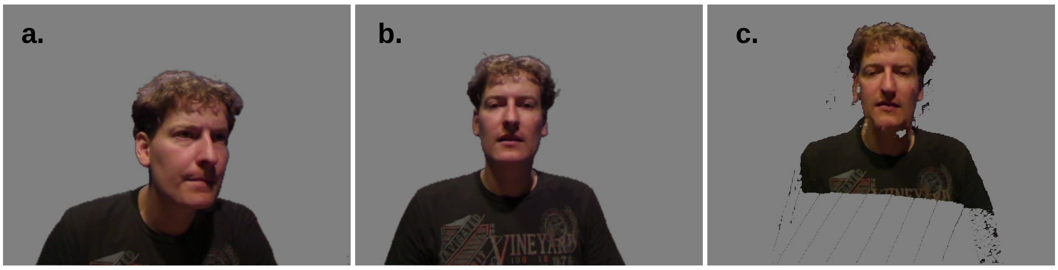


Figure 1.1: Different representations of video conversation partner: a. recordings represented without manipulation; b. intended representation towards eye contact; c. affected representation due to technical realization process.

With currently available recommendations and standards related to quality evaluation of visual representations, the quality can be differentiated by the rating of overall quality, comparisons with other representations (often with reference material), or by means of predefined attributes describing quality e.g., sharpness, color, image resolution, etc., in order to find out why the quality was rated as being good or bad. These attributes are based on already familiar distortions and on evaluators (experts and naïves) and their definitions. This quality description and rating is different from the assessment of distortions shown in the right part (c.) of Figure 1.1. Well-defined attributes related to the subjective quality perception of these errors are missing. In order to support eye contact in video communication, 3D video objects can be created out of recordings from two separate cameras, which requires considerable effort concerning the technical realization, as discussed, for example, by Weigel [2] (see also Chapter 2). A balance between the desired quality of the visual representations and the produced quality with available equipment and resources would be beneficial for both: the user and the technical realization process. For this reason, this thesis investigates mainly two issues:

- the challenges about defining quality features that describe the subjectively perceived quality.
- the question of how the technical realization process for free viewpoint video object

creation can be interlinked with the definition of resulting quality (by including the defined quality features).

Quality assessment of visual representations seeks to investigate the viewers' requirements and therefore considers human visual perception. In the context of visual representations it allows the comparison of different procedures for recording, compression, transmission, or presentation. Different methods have been developed in order to judge the resulting quality with objective measures or via subjective assessment. For this judgment, models have been developed in order to define the related notion of *quality* itself. Formalized best practices to solve specified problems have been developed in order to support the technical realization process and to make the visual representation appealing to the end user (similar to design patterns, a tool in software engineering, described, for example, by Gamma et al. [3]). In the field of audio engineering, quality taxonomies have been developed to link the technical realization process with its quality perception (see, e.g., Silzle [4], or Jekosch [5]). These activities cover special use cases and different usage contexts and outline their influence on quality perception. This work makes use of these or similar available concepts in order to connect cause and effect and to define the available relationships between produced and perceived quality in the context of eye contact support by 3DVO for video communication. Following evaluations of this concept, an exemplary definition of a quality taxonomy is investigated.

As a further stage after 3D visualizations, free viewpoint video applications enable users to navigate interactively and freely within the visual representation of a real world scene (see Kühhirt and Rittermann [6]). Applications such as free view-point choice on Digital Video Disc (DVD) or similar approaches on TV or online platforms are gaining increasing attention in the field of interactive media. The use of 3DVO representations for a more personalized and flexible television experience gets more and more attention (see, for example, Foss, Malheiro and Burguillo [7]). Three-Dimensional Video Objects (3DVOs), in the context of video communication, may offer communication support and make the conversations more social (see Mekuria et al. [8]). This can be achieved by technical solutions that help to overcome the lack of eye contact or the restricted freedom of choice regarding the viewing angle and distance to the dialog partner. These interactive activities are possible and normal in real face-to-face conversations. Several approaches realize this form of representation by using multiple views of the recorded scenes. The complex processing chain of creating a 3DVO can be handled in various forms in terms of acquisition, processing, scene representation, coding, transmission, and presentation.

The technical realization leads to a certain quality of the resulting visualization.

The question is how to define this quality and how to include knowledge about the defined quality into the development process. This work describes efforts that pay attention to the users' perceptions of these new visual representations supporting sociability and, theoretically, allowing interactivity. The goal is to define an extended quality taxonomy, in order to interlink Quality of Experience (QoE) with algorithm development. Therefore, the work considers the correlation between the used algorithm(s) and the achieved quality from a subjective quality assessment point of view. The aim of this approach is to gain further insights that may be applied to quality improvement, system adaptivity, and processing scalability in the future. The challenges are primarily the number of processing steps involved in the considered 3DVO creation, on the one hand, and the development of evaluation and measurement methods for visual quality in this context on the other hand.

This emergent field of research is influenced by several approaches in both image processing (see also, e.g., Pollefeys et al. [9], or Weigel et al. [10]) and the inclusion of subjective quality assessment for the overall quality estimation (e.g., Winkler [11], Jumisko-Pyykkö [12], Rogowitz and Goodman [13], Skorin-Kapov and Varela[14]).

Intended goals and the contribution of the author are presented in the next section. This will be followed by a description of the structure of the thesis. Section 1.2 explains the theoretical starting point and the background. In three subsections, it briefly outlines the state of the art of the topics related to the problem that will be dealt with.

1.1 Objectives and Scope

The goal of this thesis is to interlink the Quality of Experience (QoE) with the Technical Realization (TR) (see Figure 1.2). Therefore, a model is proposed that relates the factors influencing the QoE to the processing steps of the examined 3DVO creation process and various TR.

In this case, obstacles are the formulation of human perception and the weighting of the various quality influencing factors in this special context. This work aims at the definition of an interlinking model, thus leading to a measure of quality, as a best case solution, based on a taxonomy (similar to that introduced by Silzle [4]). This approach

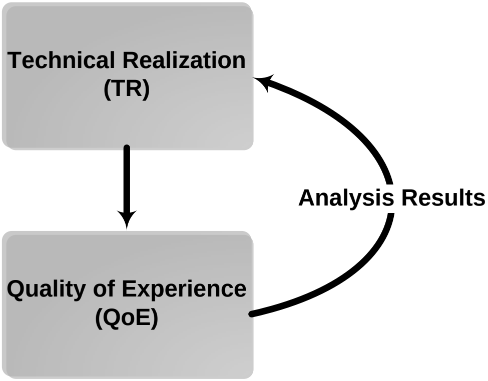


Figure 1.2: Interlinking QoE with TR.

involves a division between the perceptual domain and the physical domain, to interlink the QoE with the TR. Another intention is to define important quality influencing factors and include them in the taxonomy.

The current work uses this taxonomy approach as a starting point for the parametrization of the detected features that have a greater or lesser influence on quality. Influencing factors derived from the perceptual domain i.e., from the end users' point of view, are defined as Quality Features (perception domain) (QF). Influencing factors determined by the physical domain i.e., from the designers' and service providers' point of view, are defined as Quality Elements (service domain) (QE) (see, e.g., Jekosch [5]).

At this point, three main questions need to be answered:

- In general, the question about **ways to interlink QoE with the TR** must be answered. It is assumed that a quantization is mandatory for doing so.

- This leads to the second question, which is about **quality influencing factors and their degree of influence on both, the perceptual area and the technical realization**. However, this includes challenges based on which currently available TRs are used. Namely, these challenges are the lack of real-time functionality and, thus, the lack of interaction. Additionally, due to the development status, the test stimuli are from the low quality regime (LQR) and adequate reference material is lacking.

- Therefore, the third question deals with **evaluation methods that could be useful** for answering the second question in an adequate way, particularly in terms of the influencing features.

The mentioned challenges are addressed by paying attention to them and using a workaround (e.g., via consideration of a Wizard of Oz study addressing the lack of interactivity, as examined by Tobian [15], or by considering the influencing factor theoretically, if it is defined as for audiovideo interferences by Beerends and De Caluwe [16]). In the current work, it is assumed that the influencing elements in the physical domain are pre-defined by the used 3DVO creation process and its individual creation steps. Audio is (although it is about video communication) excluded from all these investigations within this thesis.

Based on these deliberations, this thesis focuses on the main research question which is formulated as follows:

How can the technical realization process of free viewpoint video object creation be interlinked with the definition of resulting quality and, in this way, profit from subjective quality assessment?

This question addresses two supplementary research questions, because

- it investigates the question of how to include the user's subjective quality estimation into the definition of overall quality, and

- it seeks to identify the most relevant factors that influence quality, in order to provide an extended quality model showing influences and relationships that support technical optimization.

Deriving from the context of the underlying use case video communication, several additional supplementary research questions are investigated. These questions focus on sociability support and human perception i.e., beauty, trustworthiness, honesty of conversation partner or the influence of shown background. These are described in more detail in the relevant sections: use case description in Section 2.3.2, applied evaluation and its results in Chapter 4.

1.1.1 Contribution of the Author

The contribution to the state of research on quality evaluation is constituted by the definition of QF and the definition of interrelationships to pre-defined QE, the strength of influences as well as their effects on the quality. This includes the description of the test material used, the decision about useful evaluation methods, the conduct of evaluation activities with more than 350 test participants, and the definition of an adequate model to present the findings in a way that is useful for inclusion into adaptive and scalable TR processes. To summarize, the contribution primarily addresses the following topics, which are explained briefly in Section 1.2.2:

- Explanation of 3DVO within the use case of eye contact support in video communication and, hence, the definition of QE. Note that the development of the test stimuli used in this work was realized by Christian Weigel within our project Skalalgo3d; see Section 2.3.2 and [2].
- Assessment of Quality, especially QoE and the definition of QF.
- Development of a concept for an interlinking model that benefits the TR process through knowledge about QoE and QF.

Related Publications by the Author
The following original publications are the core publications related to this thesis.

- Kepplinger, S., Subjective Quality Assessment of Free Viewpoint Video Objects. In: Adjunct Proceedings of EuroITV 2011, June 29th - July 1st 2011, Lisbon, Portugal, Doctoral Consortium, 2011, pp. 47, EuroITV 2011 Best PhD Award ([17]).
- Kepplinger, S., Weigel, C., Towards a model to interlink Quality of Experience with Algorithm Development. In: Proceedings of the International Conference on Consumer Electronics in Berlin, 06.09.2011 - 08.09.2011, ICCE-Berlin, 2011 ([18]).
- Kepplinger, S., Linking quality assessment of free-viewpoint video objects up with algorithm development. In: Proceedings of the 2012 IS&T/SPIE Conference on Electronic Imaging, Number 8293-29. IEEE Explore, January 2012. doi: 10.1117/12.906582 ([19]).

- Kepplinger, S., Profiling User Perception of Free Viewpoint Video Objects in Video Communication. In: Adjunct Proceedings of EuroITV 2012, July 4th - July 6th 2012, Berlin, Germany, Workshop: QoEMCS, 2012 ([20]).
- Kepplinger, S., Roadmap for a comprehensive Evaluation Approach on QoE of interactive and personalized TV. In: Adjunct Proceedings of EuroITV 2013, June 24th - June 26th 2012, Como, Italy, Workshop: W4, 2013 ([21]).
- Kepplinger, S., Hofmeyer, F., Gründl, M., Development of a Binocular Eye Tracking System for Quality Assessment of S3D Representations. In: Proc. of 5th Intern. Workshop on Quality of Multimedia Experience (QoMEX 2013), 3rd-5th July 2013, Klagenfurt, Austria, Industry Track, Demo, Paper-ID: D01, 2013 ([22]).
- Kepplinger, S., Tobian, D., Wizard of Oz Approach for a 3DVO Video Communication System. In: Proc. of 4th Intern. Workshop on Perceptual Quality of Systems (PQS 2013), 2nd-4th September 2013, Vienna, Austria ([23]).
- Kepplinger, S., Hottong, N., Quality evaluation of stereo 3DTV systems with open profiling of quality. In: Proceedings of the 2014 IS&T/SPIE Conference on Electronic Imaging, Number 9014-46Human Vision and Electronic Imaging XIX, 901419 (February 25, 2014), doi:10.1117/12.2038981. ([24]).
- Kepplinger, S., Linking Quality Features and Quality Elements by defining an extended 3DVO Quality Taxonomy Model. In: Proc. of 6th Intern. Workshop on Quality of Multimedia Experience (QoMEX 2014), 18rd-20th September 2014, Singapore ([25]).

These publications represent parts of the author's contribution to the manuscript and are cited at the relevant passages.

1.1.2 Structure of the Thesis

The thesis consists of three main parts:

- theoretical sections describing the related work and state of the art of 3DVO creation in Chapter 2 and quality evaluation in Chapter 3,
- practical sections focusing on the applied evaluation activities and their results in Chapter 4, and

- approach to an interlinking model that brings the QoE and TR together by using the results presented in Chapter 4, including a theoretical part focusing on different approaches to benefit the TR through subjective quality assessment in Chapter 5.

Therefore, Chapter 2 provides a definition of free viewpoint video objects and an introduction to the technical background. This includes a general description of the use case of video communication, the related project Scalable algorithms for 3DVO under consideration of subjective quality factors (Skalalgo3d) (described in Section 2.3.2), and the specific approaches chosen in this project.

Definitions concerning quality and quality influencing factors are outlined in three related sections in Chapter 3 (namely technical elements, perception-based features, and scalability), which also contains theories and definitions related to quality evaluation.

These definitions are taken up again in Chapter 4, which focuses on the notion of quality evaluation in a practical way. This includes a description about the usefulness and applicability of different evaluation methods, the methodological approach used, and the evaluation steps, in particular. The chapter outlines and discusses the results of the research.

Finally, Chapter 5 presents an extended taxonomy that shows quality influencing factors and how they can benefit prospective technical development and processing algorithms of free viewpoint video objects. Chapter 6 concludes with a summary that leads to future work.

1.2 Theoretical Starting Point

1.2.1 Background

Previous work has assessed the quality of 3DVO representations and has created a measure for describing it. Rittermann's [26] approach to creating a quality measure for 3DVO functions as a basic starting point concerning the goal of the current work. However, this approach includes subjective quality estimation only to a small extent. Although some quality influencing factors are already defined e.g., in Rittermann [26], it is expected that, under further inclusion of subjective quality estimation and under consideration of the special use case video communication, these defined factors can be complemented.

Based on the different ways by which 3DVO can be created, and on the whole process of TR (see Chapter 2), various factors become more or less important.

Existing hurdles are twofold: the various TRs and their respective 3DVO creation processes (because there are many promising approaches in use) and the definition of the right method in order to include subjective quality estimation. This asks for activities to detect (further) quality influencing factors and to include this knowledge into the development of an adaptive, effective, and scalable TR. Therefore, it is useful to create an extended taxonomy or an interlinking model (maybe leading to a mathematical measure) which is able to consider humans' perception in its most adequate manner, in order to define the representation quality of 3DVO.

1.2.2 The Related Topics at a Glance

This thesis consist of three main topics: First of all, the description of 3DVO and different approaches to TR. Second, the theoretical description of quality, QoE, and its estimation of these particular visual representations (i.e., 3DVO under investigation). The third topic is the quality taxonomy, which is seen as the prospective gain for TR by QoE inclusion. In the following, a brief introductory survey of these three areas is given. These areas will be considered in more detail in the related chapters of this thesis.

Free Viewpoint Video Object (3DVO)

In this thesis, a 3DVO is defined as a visual representation of a recorded, time variable, three-dimensional natural object e.g., human, that has its origins in natural sources (see Section 2.1). Herein, 3DVOs allow more interactivity within a video scene. This form of representation e.g., of a natural object such as an anchor woman, as shown in Figure 2.1 on page 20, theoretically allows the viewer to watch the 3DVO within a scene from different viewpoints, as described by Rittermann [26]. Another form of application is to exchange certain objects within a scene, based on certain viewer preferences or on situations as presented by Foss, Malheiro, and Burguillo [7]. In the use case of video communication, similar applications are conceivable that represent a more natural conversation environment. The special use case video communication considered in this work includes the usage of 3DVO in order to support eye contact, as investigated by Weigel and Treutner [1].

3DVO can be created in many different ways. To date, different approaches have been available for different use cases. As shown in Figure 1.3, the general processing chain consists of the acquisition of information e.g., stereo recordings via two cameras or multi view systems, the processing of the recorded information (including segmentation), scene representation e.g., more image-based techniques or geometry-based modeling, the coding e.g., predictive mesh coding, transmission e.g., via internet protocol, and presentation e.g., on a traditional television device. Current common methods of scene representation can be differentiated into more geometry-based e.g., with static texture and single geometry, and more image-based e.g., light-field rendering, 3D modeling processes. As presented in Pollefeys et al. [9], and shown in Figure 2.2 in Section 2.2 about 3DVO creation, a distinction between geometry-based and image-based 3D modeling cannot be sharp, because there are different combinations available and similar manners of description and processing. Approaches also exist that combine parts of geometry-based and image-based methods of scene representation. These methods, as well as the concepts of segmentation, extraction, and rendering are further described in Section 2.2. An overview and more detailed descriptions of the practices are given by Pollefeys et al. [27]. A description of multiple view geometry can be found in [28]. There are (still) unanswered questions about novel algorithms, especially for image analysis and synthesis. Some approaches tend to use a mixture of the described creation processes and consider actual developments in the video coding and transmission area. See Chapter 2 for further information on view acquisition, processing and scene representation. Smolic et al. [29] also considered compression and transmission in order to study a complete system for an efficient 3DVO creation and presentation process.

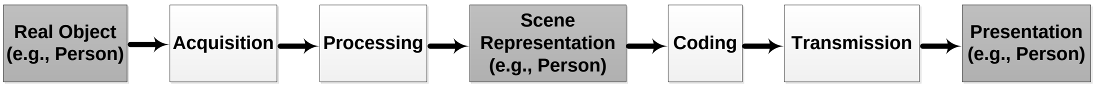


Figure 1.3: Simple overview of a 3DVO processing chain.

The 3DVO generation process considered in this thesis is used to achieve gaze correction in order to support eye contact in video communication. In Figure 1.4 the upper part (a.) shows a general architecture of an image-based 3DVO creation process and the lower part (b.) a modified process, as investigated in this work. A more detailed description of the current status of 3DVO creation processes and existing approaches is given in

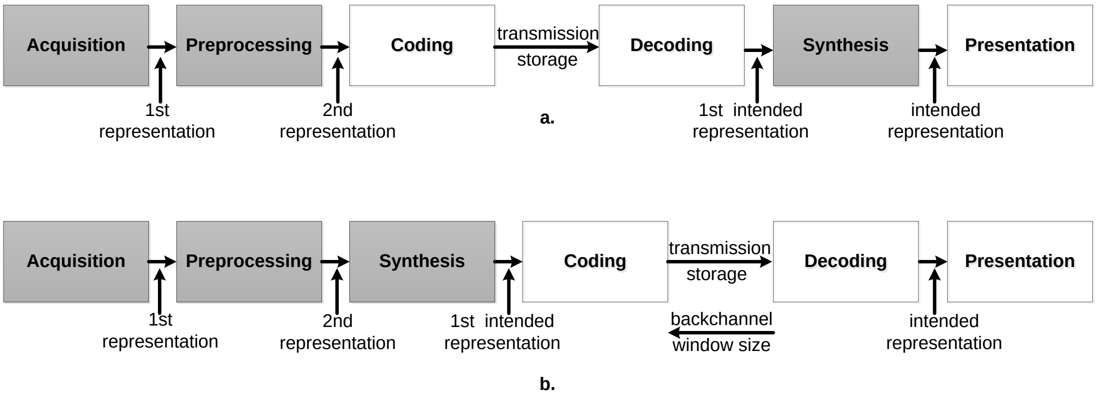


Figure 1.4: a. General model of the 3DVO creation process for image-based gaze correction. b. Modified model for video communication where synthesis already happens before transmission (after Weigel and Treutner [1]). The grey blocks are considered in this work (as introduced by Weigel [2], pp. 66).

Chapter 2. This is concluded by the description of the use case of video communication and of the TRs used in this case in Section 2.3.2.

Quality Assessment

Quality estimation guarantees a certain level of quality. Jekosch [5] defined quality itself as the results of tests and assessments. Several norms or recommendations (e.g., DIN, ISO, EBU- or ITU-Recommendations) have been developed to establish design and development outputs of high quality. But the question of how high quality is defined remains and, furthermore, how the level of quality can be estimated (see also definitions on *quality* and *influencing factors* in Chapter 3). In order to meet a high level quality assessment, various approaches exist in the area of multimedia application development. In the field of video quality estimation, the terms Quality of Service (QoS) and Quality of Experience (QoE) are common. In the field of Human-Computer Interaction (HCI) the term User Experience (UX) is common. QoE and UX intend to include the end user's subjective point of view. However, definition and usage of these established terms sometimes varies. In the following, I outline existing definitions and how the terms are used in this work.

QoS is defined as "the collective effect of service performance which determine the degree of satisfaction of a user of the service" by the International Telecommunication Union (ITU) in [30], whereas Nokia describes QoS as "the ability of the network to provide a service with an assured service level" in a white paper from 2004 [31]. Nokia also points out that it "is intrinsically a technical concept. It is measured, expressed and understood in terms of networks and network elements, which usually has little meaning to a user. QoS is a subset of the overall QoE scope. Although a better network QoS in many cases, will result in better QoE, fulfilling all traffic QoS parameters will not guarantee a satisfied user". A similar statement concerning the relationship between QoS and QoE is made in a white paper on QoE by Le Callet et al. [32]. Moreover, it is emphasized that, for this reason, QoE is highly dependent on QoS.

QoE itself is defined as "the overall acceptability of an application or service, as perceived subjectively by the end user" by the ITU in the ITU-T Recommendation P.10/G.100 [33]. This definition continues by these two notes:

- "NOTE 1: Quality of experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.).
- NOTE 2: Overall acceptability may be influenced by user expectations and context." [33].

Le Callet et al. [32] define QoE as "the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state." This quotation is taken from a "working definition" including the possibility that it may evolve further and be refined with advances in QoE research.

UX is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" in ISO 9241: Ergonomics of human system interaction: Part 210 [34]. A handful of slightly different definitions of the concept of user experience exists, of which the common thread, namely "core concepts of UX and the clarified different perspectives on UX", as outlined by Roto et al. in the UX White Paper [35], are considered to be useful definitions for this work. This includes the consideration of UX over time, which is influenced by various factors (e.g., context, user, system). For all three concepts, there exists a long list of evaluation tools and methods in order to judge QoS, QoE or UX.

Defined measures, parameters, and assessment methods are available to rate video quality objectively see, e.g., Winkler [11]. The concept *video quality* as it is used by Winkler [11] can be understood as an umbrella term. This term includes visual quality aspects focused on digital video that investigate potential quality influencing factors from the technical point of view (compression, transmission...) to the influence on subjective quality and different ways of measurement of these factors. These are, for example, pixel-based metrics such as the Mean Square Error (MSE) that measures image differences, or the Peak Signal-to-Noise Ratio (PSNR) measuring image fidelity under the usage of uncorrupted original images as reference, which are described in Section 3.3.2 on page 61. Existing approaches to subjective assessment of the quality in this context mainly focus on the evaluation of video quality in different usage contexts and scenarios e.g., laboratory vs. home environment, as mentioned in the ITU-T standard BT.500 [36]. Methods to evaluate subjective quality assessment of multimedia are being developed within the area of QoE as well as in UX research activities. Activities are being undertaken in order to clarify the different approaches to understand the QoE of new technologies as summarized in Geerts et al. [37] and the QoE White Paper [32]. However, the quality estimation of emerging 3DVO technology, particularly in video communication, has not been addressed to date (see Schreer et al. [38]). Especially the inclusion of subjective quality assessment is still underrepresented (see Rittermann [26]).

Decisions about choosing the right method depend not only on the research question(s) but also on the type of method itself, the way information is collected, whether the data should be qualitative or quantitative in nature, if the judgment is to be reached objectively or subjectively, in which development phase the assessment takes place, and who is conducting or participating in the evaluation itself. Additionally, a variety of standardized methodological approaches exists, especially in the area of perceived quality evaluation in the context of motion pictures (see Table 3.2 in Section 3.3) and for approaches with mixed method designs (see Table 3.3 in Section 3.3).

The focus of this work is on subjective quality assessment, concentrating on the QoE of 3DVO creation processes used in video communication and their interlinking to the TR (whose QoS should be estimated separately). Here, UX, mentioned above, includes several additional aspects e.g., context, pre-experiences, expectations, which may be useful to consider in a next step. However, the remaining challenge is to choose the adequate evaluation method useful for QoE assessment and to create an interlinking model.

Thus, a list of purposes and criteria has been defined that rates the suitability of an evaluation method. This is used to support reasoning and argumentation why a particular method would (not) be useful for particular requirements in this work (see Section 4.1). A more detailed description concerning existing approaches to quality estimation, and the methods actually used for quality estimation of 3DVO quality developed for video communication usage will be provided in Chapter 4. Chapter 5 describes how the evaluation results are used for an interlinking model.

Interlinking Model

Jumisko-Pyykkö et al. [39] point to the difference between produced quality and perceived quality and their different levels, leading to a minimum accepted perceived quality. The following Figure 1.5 shows the possible relationship(s) between the TR and the QoE including quality perception and quality acceptance at different produced quality levels. However, the relationships between objective and perceived quality measures for any type of video are not always linear, and the relationship might not even be monotonic. Possible effects from cognition e.g., leading to an uncanny valley (This is an empirically measurable effect that, however, appears to be paradoxical concerning the acceptance rate of artificial human-like figures. This effect is investigated in robotics research (see, e.g., McDonnell and Breidt [40]) and has to be considered when investigating the perception of virtual representations of humans.), or other influences derived from the special

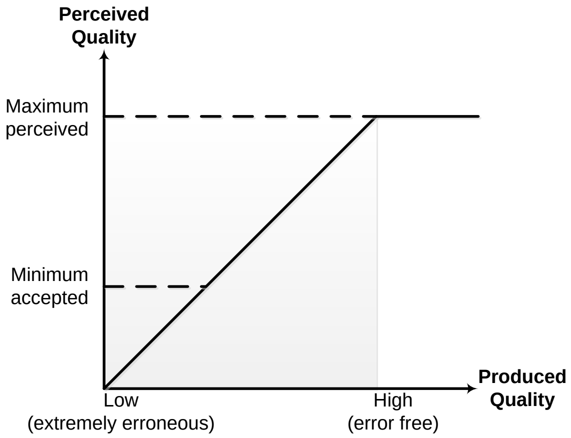


Figure 1.5: The levels of produced/perceived quality after Jumisko-Pyykkö et al. [39].

use case of video communication may create such a non-linearity. Korhonen et al. [41], for example, investigated the influence of and differences between temporal and quality scalability in video streaming applications. During experiments using low resolution video sequences, they detected that distortion has more influence on the perceived subjective quality, compared to frame rate. They also stated that the results of their quality assessments depend on the kind of content presented and on other factors that influence perception.

This possible non-linearity and non-monotony has to be considered. Therefore, a quality taxonomy is developed by taking into account possible determining influences on perceived quality and their interrelationships, as displayed by the arrows in Figure 1.6 on page 17, and combining the TR with all the QEs defining the QoS and the QoE, including its defining QFs, as used by Silzle [4], which are described more detailed in Chapter 5. This taxonomy is used as a tool in order to define a user-centered quality measure for 3DVO quality in the use case video communication.

It is regarded as advantageous for algorithm development and usage to find a way to converge QoE to algorithmic processes (the TR). The consideration of human quality judgment supports the effort of achieving technical scalability, efficiency, and adaptivity. Using a literature review, a list of theoretical concepts to meet an interlinking approach is developed (see Section 4.2.1 (B)). Approaches that bring together the TR and the quality assessment, such as norms, standards, qualifications, or rules, design patterns, guidelines, or styleguides, recommendations, taxonomies, semiotics, and heuristics, are considered (see also Kepplinger and Weigel [18]). Of these, a handful of implementable concepts (namely: norms, taxonomies, and design patterns) is described for concrete use cases in the literature. Based on this knowledge, proposal for an extended taxonomy approach to a useful concept development for the concrete use case of 3DVO creation evolves and is described in Chapter 5. This takes advantage of the concept taxonomy. Figure 1.6 displays an overview of the exemplary taxonomy application, as it is used within this work. The grey boxes represent the QF, e.g., blurriness, or color distortions, that have to be detected by subjective quality assessment. The white boxes represent the QE, which are defined by the TR used e.g., method for segmentation of recorded information, way of scene representation and the detected or defined (inter-)relationships are represented by arrows. Taking into account that the developed taxonomy not only has to outline the allocation of QE and QF to different levels and a hierarchy chart, more metric-based concepts are examined in a further step, in order to represent influences

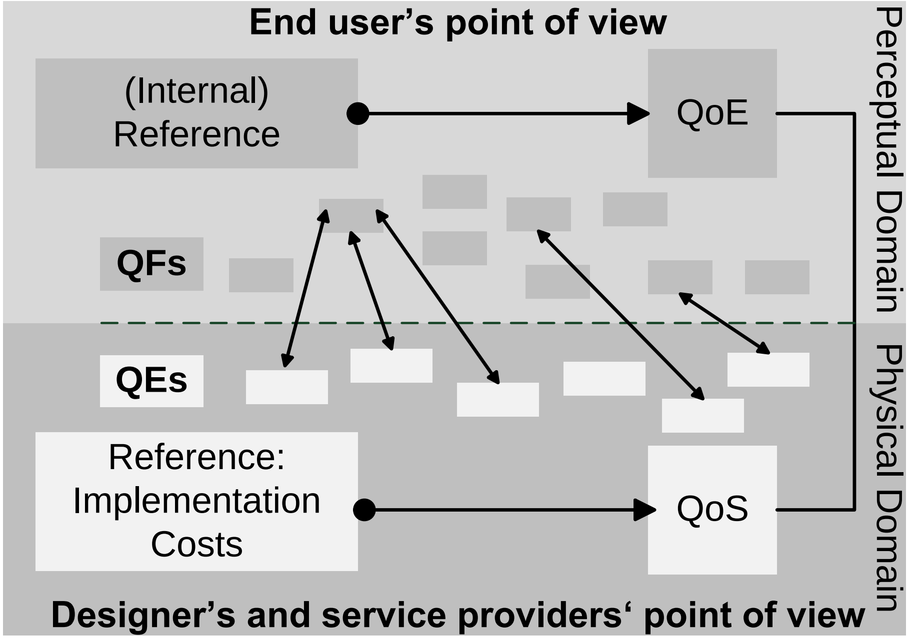


Figure 1.6: Exemplary taxonomy application on free viewpoint video, after Silzle [4].

themselves and the degree of dependencies (see Chapter 5). This includes a suggestion about how to integrate the knowledge about the perceptual domain (see also Figure 1.6) into the TR processes.During the 25th meeting of the Fernseh- und Kinotechnische Gesellschaft (Engl.: German Association for TV and Cinema Technology) (FKTG) in May 2012, Foy from Tektronix GmbH [42] presented a real-time solution for the integration of QoE measuring with a standardized set of parameters for common 2D broadcast environments. Another approach in this direction, the A/V Error Detection Libraries for different kinds of audio-visual errors, is presented by the Fraunhofer Institute for Digital Media Technology (IDMT) within the framework of their A/V Analyzing Toolbox [43]. These activities point out, additionally, the need for the current work, because knowledge about quality influencing factors and their degree of influence in the area of 3DVO usage may be very useful in the future.

2 Free Viewpoint Video Object: Theory behind creation and usage

The following section describes 3DVO and the processing chain of its creation, as introduced in Section 1.2.2, in more detail. It addresses following questions:

- What is a free viewpoint video object 3DVO?
- Which kinds of 3DVO development processes exist?
- What are they used for?

This section includes the definition of 3DVO and an introduction to the technical background concerning the creation process. The processing chain considered, and as displayed in Figure 1.4 on page 12, uses several ways for acquisition, processing, scene representation, coding, transmission, and presentation. These creation steps can be realized in a variety of ways which are outlined briefly here. After the definition, Section 2.3.2 presents the creation steps examined in the current work for the use case video communication. This thesis considers the processing chain for the creation of the test data used within the project Skalalgo3d (see also [1] or Section 2.3.2). Finally, quality influencing factors deriving from the processing chain (i.e., QE defined by the used TR) are emphasized and listed for further usage in Chapter 5 in which the development of an interlinking model is presented.

2.1 Definition

Smolic et al. [44] describe a 3DVO as follows:

> “The 3DVO is dynamic (moving and deforming over time) and provides the same functionality as conventional computer graphics models (free navigation, integration in scenes) but in contrast represents a real world object.”

In this work, a 3DVO is defined, after Rittermann [26], as a visual representation of a recorded, time variable, three-dimensional natural object e.g., human, that can be viewed by choosing the point of view interactively. It has its origins in natural sources i.e., recordings of a real world object, no Computer-Generated Imagery (CGI). Herein, natural media objects are defined as recordings from a camera or microphone (see also ISO 14496 [45]), according to [26]. Figure 2.1 shows a 3DVO, namely, the anchor woman embedded in a virtual scene.

According to [26], synthetic video objects correspond to the results of a computer animation of a 3D-object. Such computer animation results are not investigated within the evaluation activities of this work. Hence, they are only mentioned as part of the state of the art section. Rittermann [26] points out that “there are representations which are mixtures of natural and synthetic objects existing, which is possible through the usage of for example the Synthetic-Natural Hybrid Coding (SNHC)”; see also ISO 14496 [45], as mentioned in [26].

For several years, the topic 3DVO has already been part of standardization activities and standards [26]. To date, various approaches to 3DVO creation have been regarded using rather broad and general formulations referring to 3DVO as the result of some kinds of multiscopic recordings, under-representing the allowance for interaction. A 3DVO may be a representation of a dynamic real-world scene from arbitrary perspectives, like Free Viewpoint Video (FVV) [26]. However, within the project Skalalgo3d, only the creation process for 3DVO is used but interactivity is not provided to the user (see also Section 2.3.2). This means, to the user it is an ordinary 2D video representing different perspectives based on corrected positions.

Starting with information from an overview about 3D and FVV and related standardization activities in Motion Picture Expert Group (MPEG), provided by Smolic et al. [47],

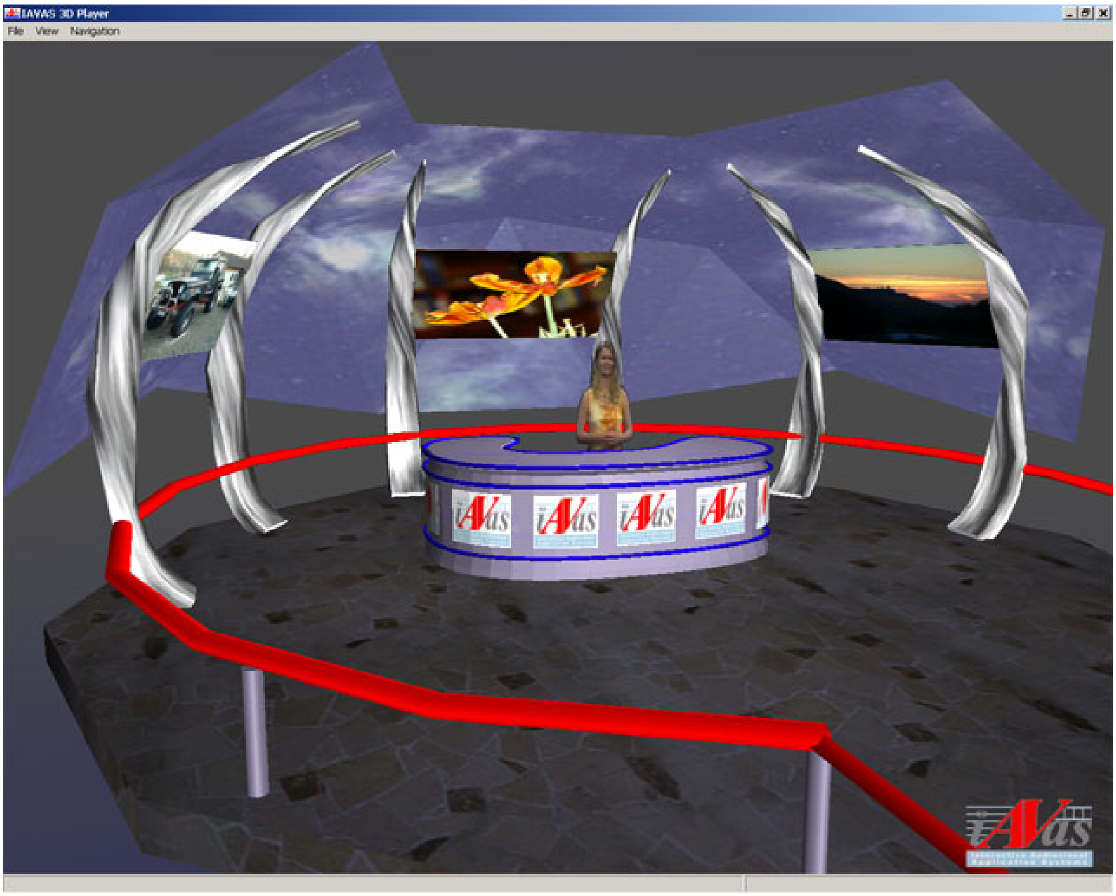


Figure 2.1: 3D Video Object created within the project IAVAS [46].

the following chronological list is given. It exemplarily presents movements and activities concerning the creation of 3DVOs, their representation processes as well as coding which shows the improvement and the diversity of various 3DVO TR processes:

- In 1998, the International Organization for Standardization (ISO) published its work about the "Generic Coding of Audio-Visual Objects, Mixture of natural and synthetic representation" within the MPEG-4 standard [45], in which media objects and their respective object- and scene descriptors are defined. These provide the possibility of (visual) object exchange and interactivity.
- In 2003, a further contribution mentioned multiscopic recording and the applications and requirements for the MPEG group formed in 2001. It works on 3D Audio-Visual (MPEG group working on) (3DAV) including the so-called experi-

ment EE2 - Free Viewpoint Video [48]. This experiment deals with multiple view coding and view generation.

- In 2005, Schreer et al. surveyed algorithms regarding 3D video communication [38]. This provides descriptions of approaches covering the whole production chain from acquisition to representation in video communication use cases.
- In 2009, Feldmann et al. provided a usage example of an immersive video conference [49]. This is important for this work, because it functions as a use case example and, therefore, as a quality reference for the intended development.
- In 2011, the ISO published a description of the MPEG-4 AFX, including 3D video objects [50]. AFX offers a toolbox including hierarchically ordered models to build synthetic MPEG-4 environments.
- In 2012, Foss et al. suggested a usage framework for introducing different 3DVO into existing videos for commercial use and to support interactivity [7]. This also leads to the usage of interactive possibilities at the service providers' side, and thus provides another viewpoint for 3DVO usage.
- In 2013, the MPEG approved new activities geared towards use cases and requirements on free-viewpoint television which include several of the 3DVO use cases mentioned above such as eye contact support in video communication systems [51].

As a result of these and similar activities within the MPEG, the subgroup of MPEG 3DAV, and others, standards were created: For example, the auxiliary video data representation developed for attaching additional information (e.g., parallax or depth values) to the individual pixels of a regular video stream [52] and Multi-view Video Coding (MVC), which provide view scalability at the bit stream level in the H.264/AVC video coding standard [53]. These activities mainly cover how 3DVO are represented as well as the coding part. Additionally, use cases evolved and several algorithms were considered. Currently, this description of 3DVO creation is a broad definition. There are many different procedures from image-based and model-based approaches in order to process 3DVO and to cover the topic interactivity and free choice of viewpoint as outlined in [26] and next section.

Although they are mentioned for the sake of completeness in the next section, coding, compression, and transmission of 3DVO do not play a further role in this thesis.

2.2 Creation

Various techniques for 3DVO creation are used within the whole processing chain (see also Figure 1.4 in Section 1.2.2 on page 12), starting with acquisition. Acquisition, as well as processing and scene representation, are investigated in the following chapters, with a focus on quality assessment. Different approaches lead to different problems and artifacts. This fact is an important issue to consider, in Chapter 5, for the definition of the physical domain (see Figure 1.6 on page 17) for an interlinking quality taxonomy model. The description of each investigated processing step contains a brief discussion of these problems. The main focus is on the problems occurring during the envisaged creation processes, as described in Section 2.3.2. This creation process includes, primarily,

- view acquisition,
- processing, and
- scene representation.

Furthermore, coding and transmission would be part of the whole process. But these do not play a role in the related project Skalalgo3d. Hence, coding and transmission are excluded for the further investigations within this work. Therefore, the end of this section contains a summary of problems arising from these creation steps. In the following, the envisaged parts of the whole processing chain are explained briefly.

In general, it is possible to differentiate between **view acquisition** through camera recordings (from recordings with one or two cameras, up to hundreds of cameras) or modeling. This includes different possibilities for the acquisition setup, from CGI, stereo camera setups, multi-view setups, through to a modular multi-camera setup e.g., for the use case football stadium. The number of cameras determines the amount of data that needs to be processed, the level of quality, or the amount of missing information, all of them leading to problems: either the processing of huge amounts of information or the missing information e.g., resulting in holes in the picture, see, e.g., 4.2 that needs to be refitted e.g., by using a filling algorithm. Depending on the following processing steps and uses of application, angle errors, incorrect optical parameters, self-occlusions, (temporal) data synchronization, position adjustment, and lighting conditions can occur during the acquisition with camera recordings. These circumstances can be more or less problematic.

The **processing** and **scene representation** depend on how acquisition is achieved. Therefore, different approaches exist. Current common methods of 3D modeling can be differentiated into more

- geometry-based,
- image-based, or
- hybrid

processes, as displayed in Figure 2.2.

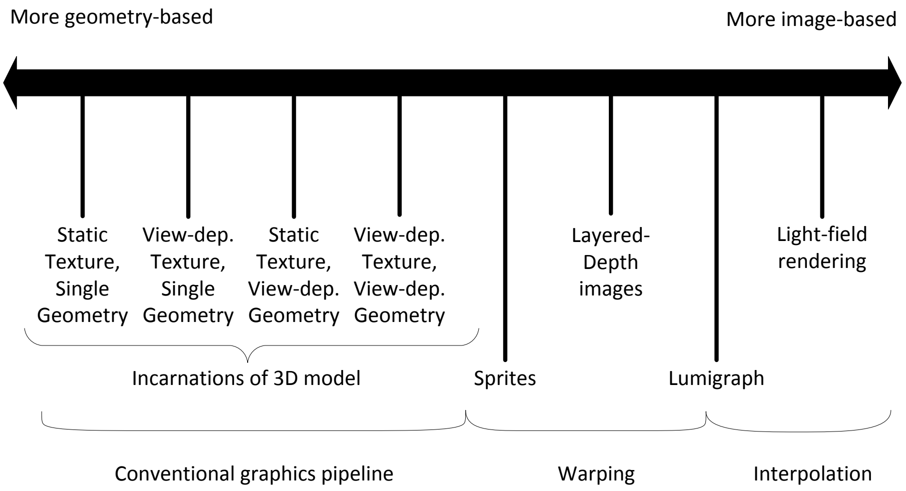


Figure 2.2: Continuum of 3D scene representations, showing different view acquisition approaches from more geometry-based to more image-based solutions (after Kang et al. in Pollefeys et al. [9], p. 164).

These different approaches to view acquisition are described as briefly as possible related to their complexity.

More geometry-based approaches

The first group of approaches uses geometry as a basis and, based on this, a specific model of a scene (single geometry) or each representative view of a scene (view-dependent geometry), including the representation of light. Thus, the parameters of this model need to be identified and textures have to be created. This necessitates the computation of the structure(s) of each scene. Model-based approaches try to reconstruct a temporally variable model of a created object, based on the principles of 3D computer graphics and animation. This leads to similar forms of usage, which may have the advantage of embedding the object into the appropriate lighting situation, including shadows. The acquisition is different from the image-based process (described below), as it uses modeling (and not recording via camera of a real object) as its basis. However, additional systems are often used to gather more information about the room and the light conditions e.g., depth cameras or light detection and ranging systems. Various techniques are used for the extraction i.e., different estimation algorithms, see also Table 2.1, reconstruction e.g., as in computer graphics, and for the texturization i.e., the description and color of the surface of the 3D model. The choice of a technique may determine its similarity to reality. More detailed information about the processes within scene extraction, about what happens during reconstruction and texture mapping and about advantages and disadvantages at each step of the different approaches can be found in [27] and in [2]. See also Section 2.3.2, explaining advantages and disadvantages of each method. The approaches investigated in this work are described in the following sections.

Table 2.1 presents different modalities of the methods used for extraction, representation, and texture mapping in the context of more geometry-based approaches. Additionally, there are procedures, such as voxel coloring, that do not separate extraction, as well as texture mapping or relief texture mapping that cannot be simply integrated into this table (see Rittermann [26]).

Problems to be dealt with in geometry-based approaches are mainly related to (see Kang [54]):

- execution time depends on the visual complexity of the scene.
- execution time relies on the hardware accelerator for speed.
- sophisticated software is required for realism.

Table 2.1: Different (more geometry-based) methods of extraction, representation, and texture mapping (based on [26], p. 42).

Extraction	Representation	Texture Mapping
Methods defining the surface by using depth maps	Dynamic 3D wire frame model	Static
Methods based on volume intersection usually using silhouette information (needs segmentation)	Voxel (organized in a hierarchical octree structure) 3D video particles/fragments	View-dependent Light field mapping (using a light-field representation for single textures)
Methods using both above-mentioned approaches and allowing texture mapping at the same time (e.g., voxel coloring)	Combinations (e.g., transformation of voxel model into wire frame model)	Unstructured lumigraph

More image-based approaches

More image-based approaches use raw image data e.g., from recordings of several cameras, taking into account the plenoptic function (see [55]), together with additional information for the view synthesis. The additional information e.g., pixel correspondences, disparity map, depth maps, light rays, silhouette, image-based visual hull, or Layered Depth Images (LDI), describes the recorded object e.g., camera position, perspectives, or information about the light and basic colors, as used for light-field rendering. By interpolation of these available data, new images are generated. Hence, this allows the generation of virtual camera recordings that combine real data from several real cameras. The image-based creation process includes several techniques, which can be divided into non-physically based image mapping, mosaicking, interpolation from dense samples, and geometrically-valid pixel reprojection [54]. The *view-synthesis* is the reconstruction of new (virtual) views from recordings of given points of views. Common methods that do this are morphing, incomplete 3D (I3D), light-field-rendering, and ray-space rendering.

Morphing is an approach to generate new views through interpolation of two existing views. Therefore, information from corresponding px of existing views is used for an analytical metamorphosis [6], p. 106.

Incomplete 3D I3D uses basic principles of 3D computer graphics to some extent and seeks to achieve effective coding. It uses object-based coding of data about texture, depth, and auxiliary data that consider the MPEG-4 video encoding technique. It selects Areas of Interest (AOI) and unifies them to a texture that includes all visible parts of the object. With support from the detected disparities, areas are chosen that are nearest to the respective camera. AOI are summarized to the primary surface. To reduce artifacts (due to occlusion), a secondary surface is created from the respective complements of the AOI. Both textures are combined to a final surface using a weighting function. The final surface defines the interpolated view [56].

The method of *light-field rendering* was introduced by Levoy and Hanrahan [57] in 1996. In this method, the input images are interpreted as 2D slices of the light field, a 4D function. This function represents "the flow of the light through unobstructed space in a static scene with fixed illumination" [57]. The authors propose a parameterization of lines by their intersections with two planes in arbitrary positions (see Figure 2.3). The coordinate system is converted onto the first plane (u, v) and onto the second plane (s, t). Connecting a point on the uv plane to a point on the st plane leads to an oriented line. The authors restrict u, v, s, and t to lie between 0 and 1 and points on each plane to lie within a convex quadrilateral. They call this representation a light slab. This light slab represents the beam of light entering one quadrilateral and exiting another

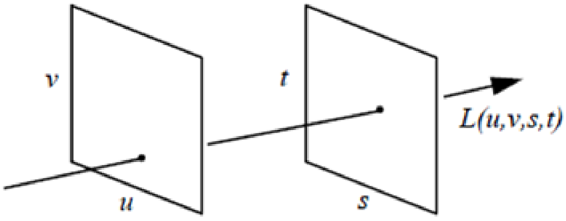


Figure 2.3: Light-field rendering: light slab from Levoy and Hanrahan [57].

quadrilateral. Integral photography and its direct and perfectly aligned acquisition, using optics, is closely related to this.

Another image-based approach similar to light-field rendering is the *ray-space rendering* that uses another representation of the light rays ensuing from a certain point in the room [9]. The ray-space solution uses a large number of cameras i.e., hundreds, for a dense capturing and transforms the information into the ray-space domain. Thus, new views are generated by selecting the useful data and inverting the transform. No 3D models or decomposition processes or any other processing steps using the geometry information are investigated.

Remaining problems to deal with in image-based approaches mainly concern:

- “Fragile algorithms that may not always find the correct correspondences in interpolating approaches” [57].
- “Restriction to regions of space free of occludes in light-field rendering” [57].
- “Challenge to fill in the gaps when previously occluded areas become visible in warping approaches” [57].
- “Sampling density to avoid blurriness in light-field rendering” [57].
- “Handling of interreflections based on changing illumination in light-field rendering” [57].
- “Lack of natural illumination and reflectance changes if the viewpoint is altered in intermediate viewpoint interpolation” [56].
- “Contour and texture errors after using a low quality disparity estimation algorithm in stereo vision” [58].
- “Dependency on input images concerning reproducible realism” [54].
- “Reliability on processor speed” [54].

Hybrid approaches

As already mentioned, different kinds of representations exist. In 2001, Pollefeys et al. [9] forecasted in a panel session on visual scene representation, that an interesting synergy between sample-based and model-based representations will occur. Besides the conceptualities 'geometry-based' and 'image-based', terminologies such as model-based representation - which is a more geography-based approach - and sample-based or px-based representation (also warping)—which is a more image-based approach, are also in circulation. Typical examples for model-based (more geography-based) representations are visual hull approaches that store the information in a mesh structure, similar to those used in computer graphics, along with textures of the scene. With respect to the sample-based approach Pollefeys et al. [9] state that "Sample-based approaches need to sample densely the degrees of freedom that are considered. This restricts the complexity of the phenomena that can practically be modeled this way." The px-based approaches use pixel-based representations such as disparity or depth maps to encode the depth of the scene. This approach will be investigated in more detail in the next sections of this work (see Section 2.3.2 on the related research project).

The 3DAV group mentioned earlier (see Smolic and Kimata [48]) investigated some of the above-mentioned approaches in terms of data representations and corresponding coding schemes. The *3D Model Reconstruction* approach uses several different ways of representation. The scene is captured with a small number of cameras (from two to thirty). On this basis, the moving foreground is segmented. This occurs in real-time using background subtraction. By using shape-from-silhouette methods that calculate an approximation of the object (the visual hull), the 3D information is reconstructed. Visual hulls can be represented in several ways (see also Smolic and Kimata [48]):

- Image-based representations (image-based visual hulls) are view-dependent and reconstruct a depth image in the desired view and shade it directly via video textures. The whole process is located in the player side.

- Polyhedral representations (polyhedral visual hulls) are view-independent and reconstruct a mesh representation of the scene by intersecting the visual cones of the object, which is rendered by view-dependent texture mapping. The known problem at this stage is the need for a dynamic mesh, which is not yet standardized.

- Point-based representations (3D video fragments) or 3D-warping are also view-independent and reconstruct an irregular point cloud, which is rendered by forward projection and image reconstruction (surface splatting).
- Volumetric representations are view-independent and reconstruct an occupancy voxel set which can be rendered by determining voxel visibility and rendering each voxel using textures.

In the following, Table 2.2 presents a summary of some of the above-mentioned methods between image-based and geometry-based approaches and also covers hybrid approaches, as outlined in [38], including different processing steps. These summaries provide an insight into squishy borders and various hybrid approaches for 3DVO processing and scene representation, which are moving towards coding and representation, as mentioned by the 3DAV group [48]. This, once again, demonstrates that there is no simple and uniquely defined, universal way for the TR of 3DVOs.

Table 2.2: Free view image generation methods and their features (adapted from [38], p. 56).

Method	Data acquisition	Data conversion	View generation
Image domain	Direct acquisition	No	Warping/ projection
Integral photography	Direct acquisition (precisely aligned)	Optical	Optical
Ray-space	Calibration and registration	Coordinate transform	Memory access/interpolation
Surface light field	Calibration and registration	Decomposition approximation	Texture mapping px-by-px multiplication
Model-based	Calibration and registration	3D model texture	Texture mapping

Similar to the processing and representation, there are several approaches available related to *coding and transmission* within the general 3DVO creation process. Smolic et al. [29] and Müller [59] provide an overview and a comparison of coding methods (e.g., predictive mesh coding, 3D mesh coding (static) (3DMC)) and its parts (e.g., representation, optimization, texture coding, reconstruction of coded scenes) of single components, based on the scene description of FVV leading to the bit stream for transmission. Various problems may occur because of the respective coding or transmission technique. Amongst others, known problems are reconstruction errors, deviations from the original images, compression errors, and the balance between data rate and acuteness of quantization [59].

Based on the application, different ways of 3DVO presentation are possible, in conjunction with an adequate input device for interaction, if this is a part of the application. In general, all systems which meet time variable visual representation requirements i.e., processing performance, demodulation, data preparation, feedback processing for 3DVO presentation are conceivable. These include, for example, TV screens or other displays (such as Personal Computer (PC) displays, tablets, mobile devices, very small screens), as well as larger screens used with a projector. Similar to the general obstacles within the scene representation and the dependency on the used method of coding or transmission, problems are (not only, but also) based on the method of presentation. This may include format adaptation, screen size, and other factors influencing the representation quality.

To summarize the current state of the art of different existing 3DVO creation processes it will be outlined that the literature (e.g., Pollefeys et al. [27], Scharstein and Szeliski [60], Carranza et al. [61], Telea [62], Rittermann [26], Cyganek and Siebert [63], Weigel et al. [10], Geiger et al. [64], Weigel and Treutner [1]) mainly considers the topics extraction, representation, and rendering methods. In their survey of scene representation technologies, Alatan et al. [65] point out that scene representation is the bridging technology between content generation, transmission, and display stages of a three-dimensional television (3DTV) system. The literature mentions coding and transmission sporadically. Some approaches tend to use a mixture of the mentioned (creation) processes with consideration of current developments within the field of video coding and transmission. Smolic et al. [29] describe coding, compression, and transmission.

In the following, the focus lies on the creation process employed in the use case video communication and in the related research project Skalalgo3d and, consequently, on the envisaged problems. These are outlined, following an introduction to 3DVO usage possibilities, in the next section.

2.3 Usage

3DVO usage may be advantageous in various areas that employ visual representations. These include medicine, architecture, the entertainment industry, as well as research and development areas, where 3D modeling of reality allows for new insights and findings (see [59]). The following list presents a selection of current available application examples, developments, and approaches:

- FVV in the video broadcast area (e.g., sports events, such as iview presented by the British Broadcasting Channel (BBC) see Grau et al. [66])
- Compositing, integration of 3DVO into (3D) movie scenes (e.g., eye gaze correction with stereovision by Yang and Zhang [67])
- LiberoVision provides options for controlling every object in the (sport) scene with the usage of the Vizrt media asset management system [68]. See also Zweites Deutsches Fernsehen (Engl.: 2nd channel of German TV broadcasting) (ZDF) [69] that provides a sports game analysis tool.
- A team at the ETH Zurich provides insights into their research on unstructured video-based rendering in which they allow interactive exploration of casually captured videos, see also SIGGRAPH video on YouTube [70].

Kühhirt and Rittermann [6] point out that 3DVO provides the possibility to allow navigation around a presented object and to freely choose the viewpoint, viewing direction, and point of time. A combination of all of these possibilities at an accordingly great expense is not always necessary, particularly not for special applications using determined changes of viewing angle for gaze correction in video communication (see Section 2.3.1).

2.3.1 Use Case Video Communication

Available investigations of a more natural video communication take several requirements into account and several methods use free viewpoint video in the context of video communication, even with more than two conversational partners (see also Schreer et al. [38]). The problem of common video communication systems is that eye contact is not supported. A further obstacle is that the user either has to look at the display to obtain information or to look into the camera in order to establish eye contact (see Figure 2.4). Eye contact is a critical factor in the fields of communication, psychology, and sociology. In 1976, Argyle and Cook [71] analyzed the role of gaze and mutual gaze in conversations and communication. Arif et al. [72] evaluate an almost real face-to-face conversation system and the most important detected shortcoming is the lack of support for eye contact.

A possibility offered by the usage of free viewpoint video objects is to support eye contact via video communication on computers, television, or mobile devices. This can be realized by way of eye adjustment, or the use of the so-called Wollaston illusion, by adjusting the displayed person's position without manipulating the eyes. Various approaches concerning eye contact support are available. In current developments for depth and position definition of the user in the room, more user friendly and cheaper set-ups are possible, as demonstrated by Kuster et al. [74], who present gaze correction for home video conferencing. Already in 1969, Stokes [75] investigated human factors that should be considered in modern picture phone station sets. Here, the users are provided with the possibility to select a desired object field, using zoom and height features for more convenience. The goal of many video conferencing system developments is and has been to support sociability and natural face-to-face-like conversations (see Chen et al. [76], Yang and Zhang [67], Eisert [77], Schreer et al. [38], Solina and Ravnik [78]).

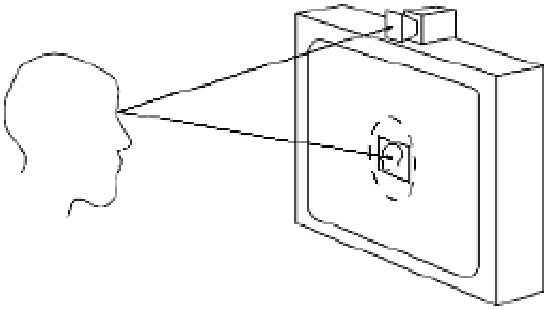


Figure 2.4: Problem of eye contact in video communication (taken from Korn [73]).

Eye contact support in video communication systems has been investigated in several ways: Murayama et al. [79] describe an approach to virtual view image synthesis for eye contact in a Television (TV) conversation system. Van Eijk et al. [80] describe the effects of gaze direction and depth on perceived eye contact and perceived gaze direction, when compared between 2D and 3D display conditions. Cohen, Colburn and Drucker [81] analyze the role of eye gaze in avatar-mediated conversational interfaces. Kollarits et al. [82] present a possible approach to technically realize eye contact via a camera/display system for videophone applications. Blue-c, as demonstrated by Gross et al. [83], is a spatially immersive display and 3D video portal for telepresence. The Fraunhofer Heinrich Hertz Institute (HHI) [84] presents an approach called Virtual Eye Contact Engine that uses high-end camera equipment and position correction of the representation. Kuster et al. [74] present an approach for gaze correction for home video conferencing. This is cheaper in terms of the basic hardware equipment and similar to the concept considered within this thesis. It differs from other approaches e.g., the Virtual Eye Contact Engine, in its usage of consumer devices (see Section 2.3.2). Several concepts involving video communication usage, as well as eye contact support and gaze correction, are also considered in more detail in student projects on video communication usage and its context (see [85], who investigates the definition of the usage context of video communication for the private conversation context as well as for the professional working area) and on the simulation of a video communication system for interaction evaluation processes. For more information concerning this, see Tobian [15] or Kepplinger and Tobian [23], who examine an approach to overcome the lack of real-time capability for evaluation activities.

2.3.2 Related Research Project Skalalgo3d

Skalalgo3d is an acronym for the research project *Scalable algorithms for 3D video objects under consideration of subjective quality factors.* This project has been conducted at the Institute of Media Technology at the Technische Universität Ilmenau (Thuringia, Germany) and has been supported for three years (2009 - 2012) by the Deutsche Forschungsgemeinschaft (Engl.: German Research Foundation) (DFG) [86].

One aim of the project was to contribute to quality improvement of 3D visualizations, in particular 3DVO. That was, in terms of optimum processing and qualitative displaying. The goal was the development of adequate algorithms as well as the definition of a model or a measure for quality, including subjective assessment, taking into account the correlation between the used algorithm and achieved quality.

The application approach considers a method to re-establish eye contact in a video communication scenario. The aim is to solve the problem of an existing misalignment between the camera and the communication window, as visualized in Figure 2.4 in the previous section. This is realized with the help of the creation of a virtual view of a camera placed at the position of this window. (See, e.g., the explanatory video available at http://youtu.be/xKjTpVlL7Sk, last viewed 13th May 2016.) This virtual view is modeled as an algorithm chain of processing steps (see Weigel and Treutner [1]).

The presented approach uses a pixel-based representation of the extracted 3D information, namely a disparity map extracted from a stereo camera setup. The method consists of several processing steps. These are visualized in Figure 2.5 and described in the following and in detail in [1]. Weigel and Treutner [1] propose a method that produces spatially and temporally consistent depth maps using cross-check based filling and motion compensated filtering. This addresses aliasing artifacts during point rendering and presents a method to enhance the virtual view by image inpainting based on robust contour warping. The chain will be considered in conjunction with the proposed interlinking model in the following sections.

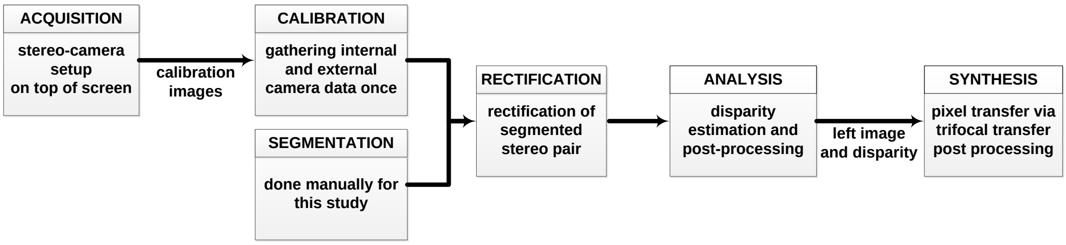


Figure 2.5: Investigated processing chain of 3DVO in video communication, after Kepplinger [19].

Anticipating the extension of the quality taxonomy used by Silzle [4] (see Chapter 6), the components of the processing steps and the modalities of the used algorithm chains form the QE. A complete description and references to the detailed descriptions of processing steps are given in [1]. Table 2.3 and 2.4 present modalities of the used processing steps for the test item creation.

Stereo Analysis:

Table 2.3: Different stereo analysis processes used for the test items applied in Sections 4.2.2 (bold) and 4.2.3 (italic).

Nr.	Disparity Estimation	Aggregation	Postprocessing
A1	**NCC**	**Gaussian 11x11**	**binary masking (Median 11x11), binary Tclean**
A2	***NCC***	***Gaussian 11x11***	***CC inpaint hole filling, binary masking (Median 11x11), hard edge cut***
A3	**NCC**	**Gaussian 11x11**	**CC inpaint hole filling, binary masking (Median 11x11), and binary Tclean**
A4	*NCC*	*Gaussian 11x11*	*CC inpaint hole filling, binary masking (MCT 16x16), Smoothing (Median 11x11)*
A5	*Mono support, Delaunay, MAP*	–	*binary masking (MCT 16x16), binary inpaint hole filling, Smoothing*

The acquisition takes place from recordings of two web cameras, which are calibrated with the usage of the Matlab Camera Calibration Toolbox (MCCT) [87], as described by Weigel [2]. The cameras use a complementary metal oxide semiconductor (CMOS) sensor of 1/3 inch (4.536 mm x 3.416 mm). The focal length is fixed at 3.7 mm. The native picture size is 1600x1200 px with 5 pictures per second maximum. The cameras use Universal Serial Bus (USB) 2.0 HighSpeed with a data rate of maximum 480 Mbit/s.

The preprocessing in the acquisition includes a segmentation to separate the foreground from the background information and a rectification in order to overcome the imperfect epipolar position of the used webcams. The source for the stereo analysis is a rectified grey scale (monochrome) from segmented stereo pairs (see also Figure 2.5). Within the preprocessing, as part of the stereo analysis, a noise reduction (Gaussian 5x5) is used for every TR.

As outlined in Table 2.3, almost all used algorithms carry out a local correspondence search with a normalized cross correlation (NCC) (which supports the reduction of outliers, even with a large window size of 21x21 px according to [2], p. 108) for disparity estimation. In contrast to the other approaches, the stereo analysis no. A5 proposes a global disparity algorithm approach, based on the work of Geiger et al. [64]. This algorithm uses a global approach with a maximum a-posteriori (MAP) estimation to define robust correspondences and uses support points for a Delaunay triangulation. For detailed information, see Weigel [2]. The majority of used stereo analysis processes (see A1-A4) employ a filter function as preprocessing with a window size of 5x5 px in order to reduce the noise caused by web cameras (see also [2]). Here, a Gaussian filter is applied on the left and on the right intensity picture before the correspondence definition upon aggregation. The postprocessing is conducted in order to exclude distortions that are still present or generated through previous processing steps, to reduce holes in the picture. This occurs with almost all used algorithms via a binary masking, using a Gaussian filter with a cost filter of window size of 11x11 px. Other algorithms use binary masking and temporal, movement-compensated smoothing (MCT) with the window size of 16x16 px. Three of the used algorithms use a cross correlation (CC) inpaint hole filling; others use a binary inpaint hole filling. Some use the binary temporal inpainting algorithm (Tclean) described in [2]. One algorithm uses a hard edge cut for distortion reduction. Another uses a (spatial) smoothing with a median filter of window size 11x11 px.

Differences between the algorithms are their use of temporal filters or spatial filters (hole filling, i.e., propagation, after Telea [62]). The purpose of temporal filters i.e., smoothing over time using a motion compensated temporal cleaning of a block size of e.g., 16x16 px, is to overcome the influence of random behavior of the collected data (see Weigel [2]) caused, for example, by the hardware e.g., thermic behavior of used sensors, which causes effects over time and does not influence quality if only single pictures are investigated.

Once the disparity map resulting from the stereo analysis is available and the preparation is done by using derectified or rectified information, together with the depth information from the left camera, the scene representation is created (see Table 2.4).

Synthesis:

Table 2.4: Synthesis processes used for the test items applied in Sections 4.2.2 (bold) and 4.2.3 (italic).

Nr.	Preparation	Virtual View Synthesis	Postprocessing
S1	**derectified information and depth information from left camera**	**Trifocal transfer**	**Fill holes (bilinear interpolation)**
S2	**derectified information and depth information from left camera**	**Trifocal transfer**	**Fill holes (bilinear interpolation, Median 3x3)**
S3	*rectified information and depth information from left camera*	*Point rendering, 5 quads*	*Fill holes (bilinear interpolation, include hc)*
S4	*rectified information and depth information from left camera*	*Point rendering, 5 quads*	*Inpaint (contour-based, include hc)*
S5	*rectified information and depth information from left camera*	*Point rendering, 30 smooth quads*	*Inpaint (contour-based, include hc)*

In doing so, either a trifocal transfer or a point based rendering, either with 5 quads or 30 smooth quads, as described in [2]), is used for the virtual view synthesis, as outlined in Table 2.4. Inpainting, based on Telea [62], was used for the post processing of synthesis no. S4 and S5. This approach is contour-based. The other algorithms filled holes via bilinear interpolation. Hence, one uses a median filter with a window size of 3x3 px. Synthesis processes no. S3, S4, and S5 included a Horizontal correction (hc).

The problems dealt with in this point-based approach, as described in Weigel [2], mainly relate to:

- artifacts outside the participants contour, e.g., caused by wrong disparities

- holes due to dis-occlusion and wrong disparities
- aliasing artifacts during rendering and re-projection to the virtual camera view (this is a problem with all px-based methods)
- flickering of the synthesized sequence due to spatial and temporal inconsistencies and errors
- unnatural position of the virtual view (flying) due to missing image information in the lower border region
- strong noise due to the low-end consumer cameras
- USB is stressed to its limits, even if comparatively small resolution (typical video chat resolution of 640x480 px) is used for capturing four streams.

This work addresses the question about which problems are envisaged within the 3DVO creation process. Therefore, Table 2.5 presents a collection of possible problems associated with the respective way of creation (see Table 2.3 and 2.4) per processing step. It summarizes the mentioned problems appearing with the stereo analysis and synthesis method. This also considers the acquisition and presentation, but not coding and transmission.

Stereo Analysis and Synthesis:

Table 2.5: Problems per processing step in the methods used for the test items applied in Sections 4.2.2 (bold) and 4.2.3 (italic).

No.	Problems within processing
A1	**New artifacts and blurriness based on Tclean**
A2	***Contortions caused by incorrect filling***
A3	**Contortions caused by incorrect filling**
A4	*Contortions caused by incorrect filling* *Aliasing caused by filling* *Blurriness caused by smoothing*
A5	*Contortions caused by incorrect filling* *Aliasing caused by filling* *Blurriness caused by smoothing*
S1	**Contortions caused by incorrect filling** **Aliasing caused by filling** **Holes and superposition through image-based rendering**
S2	**Contortions caused by incorrect filling** **Aliasing caused by filling** **Holes and superposition through image-based rendering**
S3	*Contortions caused by incorrect filling* *Aliasing caused by filling*
S4	*Dis-occlusions by resampling*
S5	*Dis-occlusions by resampling*

For the latter two TR conditions, the same conditions are true for every processing algorithm, thus the following described problems within acquisition and presentation may appear every time. Problems to be dealt with during the acquisition mainly concern:

- optics-based non-linear contortions i.e., change of objects position
- camera-induced correspondence problems
- scene-dependent correspondence problems e.g., dis-occlusions because the distance between camera and scene is too small
- synchronization problems when combining the recording of the two cameras
- mistakes within image calibration
- image loss by bus transfer or recording processes
- holes because of dis-occlusions based on camera position

Problems within the presentation of the generated representation may include problems from coding and transmission as well as influences deriving from the representation device.

Several similar problems may occur that arise from different sources. For example, spatial sampling that is too small leads to aliasing, an incorrect sensor size and the quantization factor lead to noise, color distortions may be caused by the interpolation method, correspondence costs may be too high, a foreground fattening effect may arise from the pre-production i.e., block matching, incorrectly allocated depth cues lead to jumps within moving pictures, incorrect pixel transformation leads to holes, and contortions are caused by incorrect disparities.

It is not always easy to define which problem is caused by which processing step, because the processing methods as well as the algorithms are complex and depend on different ways of TR.

This summary of problems is based on Weigel [2], the above-mentioned literature, and based on results of expert interviews (see Section 4.2.1). These problems are further investigated in the next chapters, focusing on quality and subjective quality assessment, in order to develop an extended quality taxonomy.

2.4 Summary - 3DVO and its Applications

The processing chain of 3DVO creation, as described in this chapter, is complex. Different ways of creating a 3DVO for different use cases give rise to several different kinds of quality influencing factors. Within the context of 3DVO generation for eye contact support in video communication systems, a px-based approach is used and described in the related section about the project Skalalgo3d [86] finished in 2012. This includes the introduction to the creation steps examined in this work.

The current thesis outlines the processing chain for the creation of the test data used within the project. The creation steps include different methods to overcome problems such as holes or dislocated pixels. These methods can solve the mentioned problems, but can create further problems as well, as described above. These problems and quality influencing factors within the methods used for each processing step are emphasized and listed in Table 2.5 for further usage in Chapter 5, to develop an interlinking quality describing model.

3 Theory about Quality, Quality Influencing Factors, and Quality Evaluation

Based on the definition of 3DVO and an introduction into the technical background, Chapter 3 defines quality and quality influencing factors within the context of 3DVO quality. In general, this chapter answers questions about

- how 3DVO quality is defined in this work and
- how influences are defined that may change the 3DVO quality.

This is outlined in the related sections addressing quality itself and technical and perception-based factors. These topics will be taken up again in Chapter 4, focusing on the evaluation of quality.

3.1 Quality

The ISO standard "‘Quality management systems - Fundamentals and vocabulary (9000)"’ [88] defines the term *quality* as the “degree to which a set of inherent characteristics fulfills requirements”. However, based on the context of use, quality may also include characteristics other than inherent ones, for example, beauty or taste. Section 1.2.2 introduces a definition presented by Jekosch [5]. She defines quality as the results of tests and assessments. These assessments consider a variety of quality influencing factors such as e.g., user, usage context, or application. Furthermore, concepts such as QoS, QoE, and UX as well as their relationships are introduced in Section 1.2.2. Based on these, a more detailed view on *quality* in the area of visual representations is given in this section, ranging from basic concepts, via still-image quality, through to moving pictures, stereoscopic 3DTV images, and, finally, to 3DVO quality as the focus. What is the best practice for defining 3DVO quality? Which of the existing approaches defining quality are useful? These questions are investigated in the following.

Ludewig and Lichter [89] present a quality tree (see Figure 3.1) to illustrate various components leading to quality, from a software engineering point of view. The division into process quality and product quality provides an overview of components that are important for defining QoS. This is useful for this work in relation to the definition of the QoS of the application defined in Section 2.3.1. Comparing this approach to the quality taxonomy approach as introduced in Section 1.2.2 and shown in Figure 1.6), reveals that there are several ways to define quality and different approaches to how to break down the overall quality to quality defining factors that are necessary in terms of the QoE. QoE is usually defined by quality assessment. Zacharov [90] provides a statistical model by presenting a quality defining formula based on quality assessment:

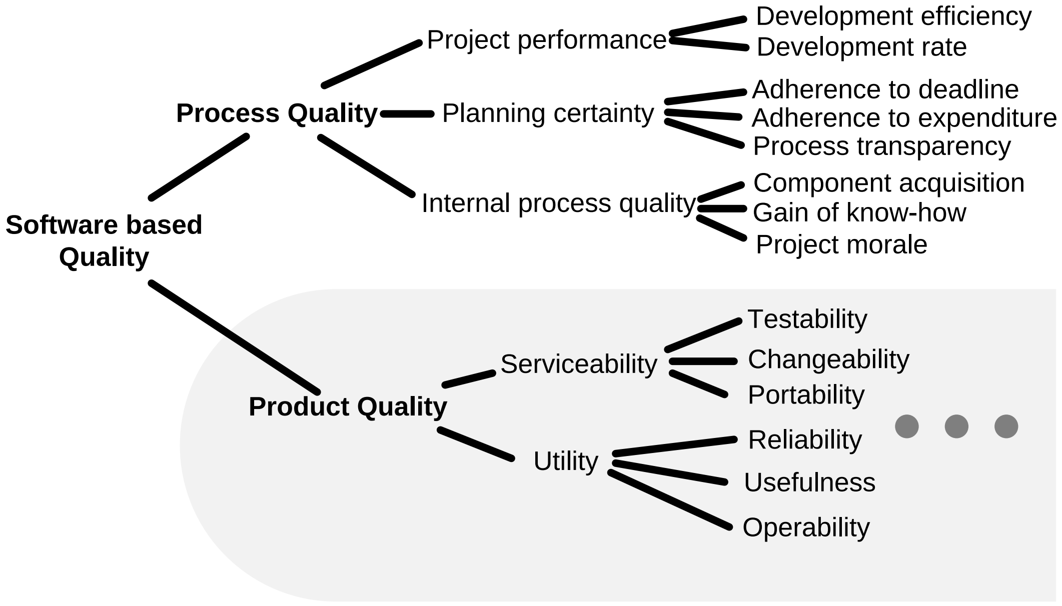


Figure 3.1: Quality Tree, after Ludewig and Lichter [89] (translated).

$$Y_{t,i} = \mu + \alpha_t + \epsilon_{t,i} \tag{3.1}$$

This formula contains:

- $Y_{t,i}$: is the i-th rating provided by a subject for the t-th stimulus
- μ: is the general mean for all stimuli
- α_t: is the effect caused by the t-th stimulus when corrected for the overall mean
- $\epsilon_{t,i}$: is the effect caused by the random experimental error

and explains it as:

- Explanatory variables:
  - Dependent variables ($Y_{t,i}$ - assessor rating)
  - Independent variables (α_t - test objects)
- Controlled variables (μ, $\epsilon_{t,i}$)
- Disturbing variables (μ, α_t, $\epsilon_{t,i}$)
- Randomized variables ($\epsilon_{t,i}$)

These concepts have in common that, in order to define quality, a *quality formation process* occurs, which includes 1) perception and 2) judgment based on a reference (either internally or on a clearly outlined reference), which leads to 3) quality rating and/or description (see also Qualinet White Paper [32]). However, in the modern multimedia representations complex, uncontrollable variables and several degrees of freedom e.g., contextual factors, influence the results. But how should these influences be considered?

Hands [91] introduced a basic multimedia quality model, in order to predict quality rating. Here, the basis is content dependency and, especially, the interrelationship between several sensory channels, mainly related to auditory and visual perception. The model itself suggests a weighting structure, following multimodal combination rules. However, Hands [91] points out that the definition of multimodal combination rules still has to be investigated, in particular by considering different tasks.

The present work focuses predominantly on visual representation. De Ridder and Endrikhovski [92] define image quality as a weighted sum of three constraints, namely, Fidelity, Usefulness, and Naturalness (FUN), and they categorize images into this FUN-space. Because usage tasks and developments may change over time and lead to a different quality influenced by several factors, the relationships between several degrees of influence have to be considered, which is possible with the FUN approach. The weighted sum of the constraints fidelity, usefulness, and naturalness defines the image quality and is dependent on task, image content, context, etc.

Engeldrum [93] defined the image quality circle, as shown in Figure 3.2, and describes image quality as "the integrated perception of the overall degree of excellence of an image". The image quality circle contains the items system/image models, visual algorithms, and image quality models, which are connected with links that allow movement back and forth between the circle's elements. The image quality circle is defined this way in order to support the prediction of possible changes in quality after alterations in technical factors, or in order to define certain technical factors necessary for achieving a specific level of quality. Within the usage of a more complex TR, as well as complex forms of usage, the definition of quality has to consider further influencing factors, like interactivity or evolvement over time.

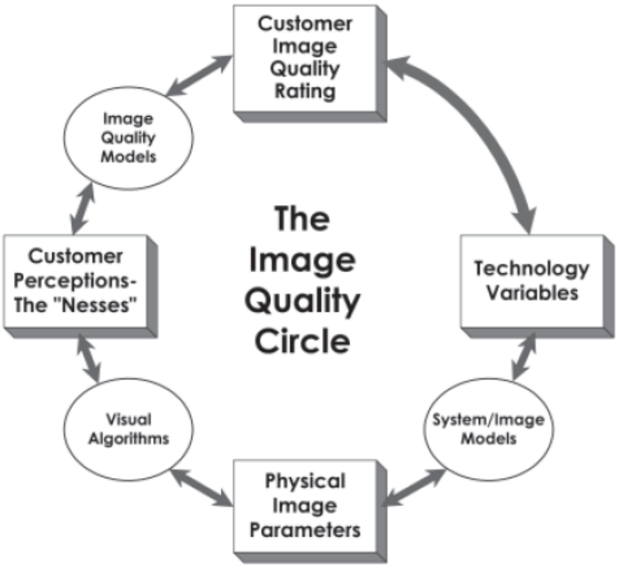


Figure 3.2: Image Quality Circle, after Engeldrum [93].

Fechter [94] defined the quality of compressed image sequences as a complex sum of different influences (linear and non-linear), based on different thresholds, which have to be fixed for every single context of use and cannot be generalized. However, predictability and defining relationships between QoE and TR is outlined as clearly as possible and is useful for the consideration of diversification and task. The entity is the user perceiving quality.

3.1.1 Quality Models

Before the next sections consider quality influencing factors in more details, e.g., whether they are QFs or QEs, this section deals with different QoE models. These models reveal different ways of how to consider quality and quality influencing factors.

With respect to the classification of factors that influence QoE, Jumisko-Pyykkö [95] defined a model of user-centered QoE. This model, shown in Figure 3.3, contains four main components: the user, the system, the context of use, and the experiential dimension. This figure excludes the example of descriptive attributes within the experiential dimensions, as given in the original figure that includes characteristics of the user, the system or service, the context of use, and experiential dimensions e.g., descriptive attributes for the investigated system or service. Jumisko-Pyykkö gives this example of descriptive attributes for mobile 2D/3D TV in the context of use ([95], p. 64). The advantage of this model is that it extends the concept of *system* and *user* as quality defining entities with the context of use and, furthermore, with empirical issues. The disadvantage is that further degrees of freedom have to be considered or controlled. For this thesis, it supports the necessity to consider the knowledge of possible further quality influencing factors based on entities other than the pure physical or perceptual domain. Furthermore, changes in the quality assessment adapted to such circumstances have to be considered.

Laghari et al. [96] provide a comparison of QoE models in a communication ecosystem, presented in Figure 3.4. There, distinctions are made, based on whether the respective model takes into account several factors from the human domain i.e., human roles and human demographic attributes, subjective QoE factor, objective QoE factor, and the technological domain, the contextual domain, and the business domain. Based on this survey of key QoE models, they propose a further QoE model and present it in a high-level diagram for a QoE interaction model in a communication ecosystem. This model contains the domains human, context, technology, and business. Here, the human has interdomain interactions with all other domains, technology to business, and context to

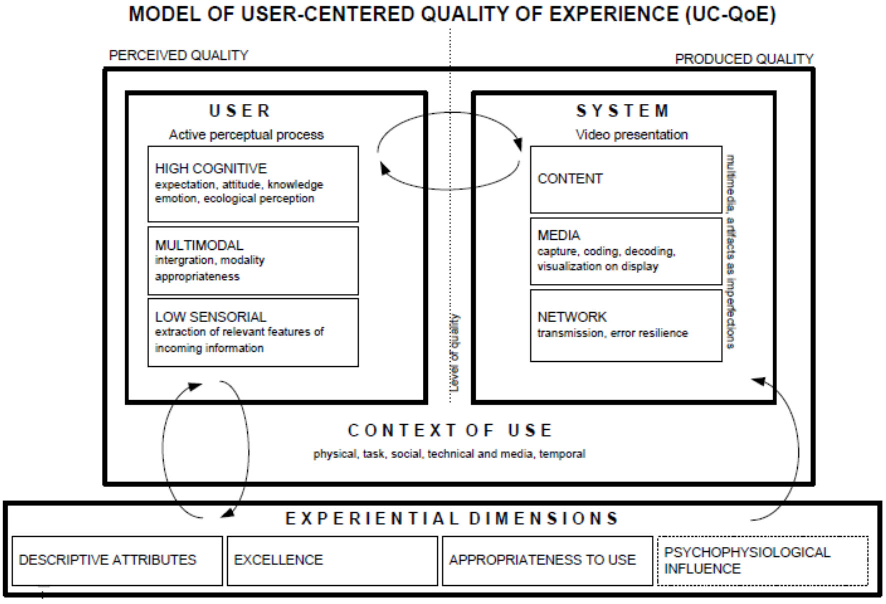
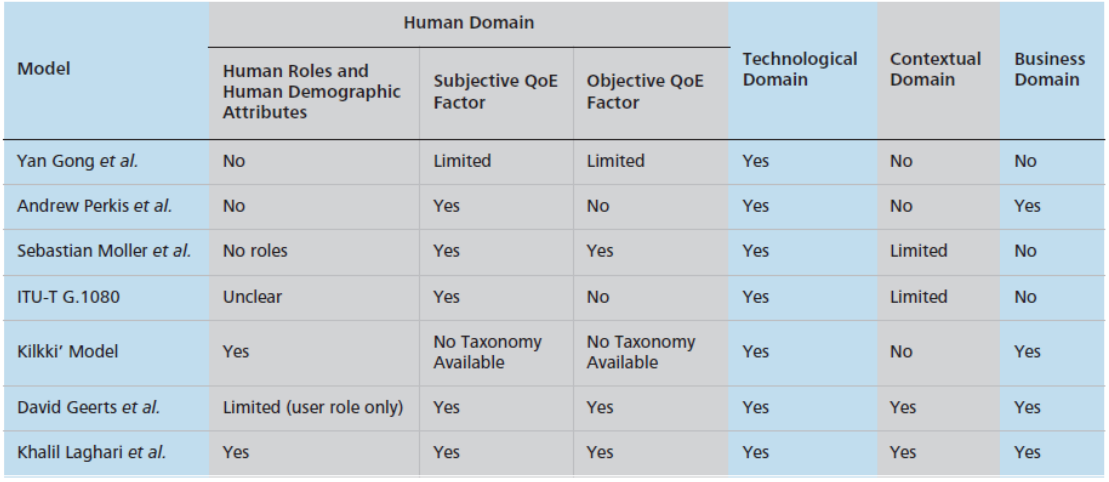


Figure 3.3: Model of user-centered QoE (except descriptive attributes) by Jumisko-Pyykkö [95].

techno-business. Laghari et al. define QoE as a "blueprint of all human subjective and objective quality needs and experiences arising from the interaction of a person with technology and with business entities in a particular context". The advantage of this model is that it surpasses existing models because of its relatively detailed definition of possible interdomain interactions. The disadvantage is that examining the cross-domain relationships in such a holistic approach is expensive and maybe more useful for already well elaborated areas i.e., where several components of the model are already known and can be pre-specified. However, for this thesis, it supports the interlinking approach and the consideration of interdomain interactions. Another model for QoE, presented by Skorin-Kapov and Varela [14], uses a generic approach in order to group factors affecting QoE. They are organized into four spaces, the Application Resource Context User (ARCU) model. The advantage of this model is that it organizes influencing factors beyond QoS in a multi-dimensional way. The influencing factors are categorized into related spaces, namely application, resource, context, and user. Furthermore, points from these spaces are mapped to a multi-dimensional QoE space including qualitative and quantitative QoE metrics. The disadvantage is that, to date, this only provides a basis for systematically categorizing influencing factors and for understanding relationships,

]



Model	Human Domain			Technological Domain	Contextual Domain	Business Domain
	Human Roles and Human Demographic Attributes	Subjective QoE Factor	Objective QoE Factor			
Yan Gong *et al.*	No	Limited	Limited	Yes	No	No
Andrew Perkis *et al.*	No	Yes	No	Yes	No	Yes
Sebastian Moller *et al.*	No roles	Yes	Yes	Yes	Limited	No
ITU-T G.1080	Unclear	Yes	No	Yes	Limited	No
Kilkki' Model	Yes	No Taxonomy Available	No Taxonomy Available	Yes	No	Yes
David Geerts *et al.*	Limited (user role only)	Yes	Yes	Yes	Yes	Yes
Khalil Laghari *et al.*	Yes	Yes	Yes	Yes	Yes	Yes

Figure 3.4: Comparison of QoE models, from Laghari et al. [96].

but the identification of these influencing factors is not considered. For this thesis, it supports the idea of using weighting factors, which are assigned to individual influencing parts (in this model: QoE dimensions), in order to develop an overall measure defining quality.

These models classify various influencing factors, in order to define quality. They have in common that all the defined factors are seen as influences for the overall quality.

Although all these models focus on QoE, the considered affecting, influencing, and severe factors are both, QEs deriving from the physical part and QFs deriving from the perceptual part. Additionally, these factors influencing quality are sometimes in between QEs and QFs. They are represented by entities such as context, business domain, or experiential factors. And they are considered differently e.g., with pre-defined interdomain interactions. The concept that focuses more on the subjective assessment is named QoE. Diepold [97] discusses the difficulties of differentiating QoE from other concepts of quality, as well as the challenges of defining quality itself.

3.1.2 Summary - Quality

The following work addresses QoE as part of the perceptual domain, including the physical domain and QoS, and therefore considers the QEs as part of this latter domain, as outlined by Silzle [4]. However, the context of use and the experiential dimensions introduced by Jumisko-Pyykkö [95], or the business domain revisited by Laghari et al. [96], are seen as crucial factors that need to be investigated.

The definitions outlined above form the notion of the concept “quality”, from image fidelity to more complex applications. However, these influence the definition of quality, as it is used in connection with 3DVO evaluation in this work, which is focused on visual representations created with a defined creation process (see Section 2.3.2). In particular, Jekosch’s [5] concept of quality, as outlined in Section 1.2.2, is considered, taking into account constructs of quality in vision. The current work uses this definition and (as addendum) the definition of image quality as given by Roufs and Bouma, indicating that “image quality as a general notion relates to both elementary and complex visual functions” [98].

Hence, the goal of this thesis is to investigate the overall quality by defining the perception-based (QF), based on givens of technical (QE) on the physical side. Based on this definition of QF, the definition of a quality taxonomy model is intended (see Chapter 5) in order to linking QEs and QFs. Thus, this concept of quality allows for the integration of several different quality influencing factors, whether they derive from the physical (QoS) side or from the perceptual (QoE) side.

Assessment methods allow the evaluation of whether quality (either overall quality or a particular part of quality, i.e., QoS, QoE, or a single factor) is *good* or *bad*, and descriptive approaches support the detection of *why* quality is rated as good or bad and based on which factors. Therefore, different evaluation methods are considered in terms of whether they were originally developed to investigate overall quality, or to focus on QoE, or on QoS (see Section 3.3). This rating and quality description considers quality influencing factors as described in the next section. The evaluation part in this thesis focuses on the definition of QFs (see Chapter 4). Therefore, QFs are formulated in Section 3.2.2.

3.2 Quality Influencing Factors

Based on the definition of the notion of quality, this section defines quality influencing factors outlined in two related sections: technical factors, and perception-based factors.

As proposed by Coolican (see [99], p.77), independent variables are also known as factors in an experiment, when discussing more complex (e.g., multi-factorial) designs. Independent variables are the conditions manipulated by the experimenter (e.g., the kind of algorithm used) in order to evaluate their effect on the dependent variable (e.g., quality rating). However, it is not always possible to control every influencing factor other than the pre-defined independent variables. Jekosch [5] compiled facts about the term "*quality*" itself and discussed different influencing factors, such as:

- expectation controlling the quality event,
- contents, function, and form of the object being assessed,
- the way the entity is being viewed,
- the way in which assessment and perception occur,
- various considerations within the assessment process, and
- the point of view about quality.

Within this "framework for a global understanding of the term 'quality' " [5], influencing factors are divided into QEs and QFs. Whereas a QE is "the foundation on which the design of an entity is based"[5], the QF is "the result of having analyzed the perceived, designed entity and which leads to a description of quality"[5]. Based on this definition of quality influencing factors, the next Section (3.2.1) focuses on the quality influencing factors derived from the used TRs, which are defined as QEs. Subsequently, the perception-based factors are investigated (see Section 3.2.2); they will be the focus of the next chapters on evaluation and the reporting of analysis results, which are defined as QFs. The advantage of this model of quality and its components (QFs, QEs) is that it allows the investigation of the overall quality from different angles and in a formalized, neatly arranged way. This approach is open to further findings. The disadvantage is that not all influencing components are usually known or clear. For this thesis, it supports the investigation of links of pre-defined QEs with detected QFs in a formalized and open way.

3.2.1 Technical Factors - Quality Elements (QE)

Technical factors are based on general functionality and functional principles. It is possible to differentiate between the duration of processes and the efficiency, including costs and effort needed to obtain a particular result. Jekosch [5] quotes the definition of the *element of quality* by the Deutsches Institut für Normung (Engl.: German Institute for Standardization) (DIN) 55350-11 [100] as the

> "contribution to the quality
>
> - of a material or immaterial product as the result of an action/activity or a process in one of the planning, execution or usage phases
> - of an action or of a process as the result of an element in the course of this action of process."

Here, quality factors are defined as nameable categories of quality elements. Defining the basis of QE as processing steps and taking into consideration their definition via cost, performance, and/or efficiency, a link to the topic scalability is established. This investigates the possibility of achieving the same, or nearly the same, level of quality whilst being more efficient or producing less costs. QEs, as they are defined in this work, are defined by the TR, in order to create 3DVO as described in Section 2.3.2 and the list of QEs presented there. This means the investigation of different stereo analysis and synthesis processes, see Tables 2.3 and 2.4. Similarly, problems of these methods per processing step, which are considered as influences on quality, and may be seen as QFs, are emphasized and listed in Table 2.5. In this concept, *resolution* or *sharpness* are clear results of QE (e.g., deriving from the kind of image recording), which can be perception-based factors as well and—if described or named identically—be dealt with as QF. Hence, the resulting quality influencing factor depends on the origin as well as the way in which it is described and differentiated.

3.2.2 Perception-based Factors - Quality Features (QF)

Perception-based factors are based on human judgments and perceptual processes. This involves underlying theoretical concepts of perception itself (e.g., Goldstein [101]) and environmental influences (e.g., Neisser [102]), the interrelation between different sensory channels (e.g., Hands [91]), or influences from interaction and various tasks (e.g., Rogowitz and Goodman [13], or Möller et al. [103]). Jumisko-Pyykkö [95] and Strohmeier [104] explain the complexity of defining individual parameters in the perceptual domain. Strohmeier explains that human judgments of different aspects of the evaluated system or service are the basis for subjective quality evaluations. Furthermore, this judgment is based on perceptual processes that include low-level sensory and high-level cognitive processing, among them knowledge, emotions, attitudes, and expectations (see [104]). Furthermore, he also describes the current actions that are being taken in order to include aspects such as contextual behavior, internal preferences, and impressions (see also the QoE models described above). However, these actions seem to be limited, because reliable instruments for tackling the descriptive characteristics of quality and/or quality preferences and the related descriptions are lacking. The constraints that Strohmeier mentions are accuracy, complexity, required type of assessors, uni-modal evaluations, or their emphasis on qualitative methods. Baird [105] defines human judgment in the visual context as a composition of subjective impression, judgment, judgment strategies induced by experimental procedure, as well as contextual effects. He also states that image quality is a subjective impression. The human observer is seen as the benchmark for quality definition, not only in the context of QFs definition (see Winkler [11]). Several activities are and were undertaken to understand the cognitive issues of visual perception (see, e.g., Neisser [102], Hollier and Voelcker [106], Hollier et al. [107], Engeldrum [108], de Ridder [109], Goldstein [101], Grill-Spector and Malach [110], Winkler [11], or Chen et al. [111]) and to include these processes into the notion of overall quality (e.g., Engeldrum [93], Roufs and Bouma [98], Rogowitz and Goodman [13]).

For perception, the physical components of the eyes are essential conditions, just like cognition, which includes all the information about learning and influences which are related to pre-experiences, e.g., context, emotions, or other influences. The visual perceptual process is often divided schematically into four parts: (a) the visual stimulus itself that affects (b) the physical eye leading to (c) the sensation when the stimulus is transformed into electrical potentials for the information processing through neurons, which finally trigger (d) the perception which includes the individual, subjective processing of all collected information and their interrelations. Establishing the notion of *psychophysics*, Gustav Theodor Fechner (1801-1887) studied relationships following a set pattern between physical stimuli and subjective perception in order to define thresholds and theories about relationships. However, this simple notation does not capture the cyclic and complex process of perception (see Goldstein [101]). Considering this, the classification of perceptual processes into *top-down*, which especially includes prior knowledge, and *bottom-up*, which focuses on the analysis and processing of stimulus characteristics, allows investigators to deal with influences from the brain itself and cognitive processes. Thus, none of the processes are regarded individually.

With the knowledge about the physical principles of the eye as well as perception and cognition for imaging technology, limitations and phenomena of vision that are relevant for the envisaged use case are considered (as described in the following sections). Based on the task of defining interrelationships, limitations and individual cognitional issues play a more or less relevant role (see Rogowitz and Goodman [13]). Therefore, the complex and broad topic of visual perception is limited in this work by narrowing it down to emerging imaging technologies in this Section (3.2.2), especially in the subsections on the use case 3DVO and characteristics related to video communication.

Hence, for this work, relevant QFs are defined based on the considered use case. Therefore, factors known from the area of visual perception (see Section 3.2.2) and from the application of 3DVO (see Section 3.2.2) and video communication (see Section 3.2.2) are taken into account when defining relevant QFs.

Perception and Cognition of Visualization Technology

Based on the physiological functions that allow vision, the phenomena of visual perception make the definition of vision more complex:

> “Seeing is believing.” (English proverb)

Beyond psychology and within the evaluation of images and videos, the visual perception and cognition has received considerable attention in engineering (e.g., Kayser [112], Fechner [94], De Ridder and Endrikhovski [92], Winkler [11], Rogowitz and Goodman [13]). The most important phenomena relevant for this particular work are *human sensory reception*, including isomorphic subjectivity e.g., as described by J. Bruner in Balcetis and Lassiter [113], and subjective quality factors e.g., as discussed by Winkler [11], p. 48, such as individual interests and expectations, display affinity, viewing conditions, fidelity of the reproduction, and interrelationship with other senses, like the sense of hearing e.g., see [114]. *Human sensory reception* is described in the Encyclopedia Britannica: “human sensory reception, means by which humans react to changes in external and internal environments” [115]. Influences in terms of visual perception and cognition could be, for example, the pre-experience and expectation of the user, which may lead to QFs that are unnatural (or natural). However, this is only one example out of many. Additionally, these influences all depend on the particular use case which may lead to effects appearing paradoxically e.g., uncanny valley as known from the research in robotics. Use case-related characteristics are described in the following, focusing on 3DVO and on video communication.

3DVO-related Characteristics

Little experience is available about the perception of 3DVO representations, in part due to its novelty and, thus, its underrepresented distribution and usage in day-to-day-life and, in part, to its complexity in terms of TR and usage context. Rittermann [26] discusses the challenges of defining cognitive issues, because these may result from 2D and/or 3D vision, as yet uncollected usage experiences, and/or from more modern visualization possibilities including Augmented Reality (AR). Despite these issues, he [26] already defined several quality influencing factors, namely, incorrect resizing, defective angles of view, occlusions, or distorted silhouettes, all of which were detected within evaluation activities when considering subjective quality assessment. As such, in the context of 3DVO quality assessment, preliminary defined influences on visual representation quality such as the ones mentioned by Rittermann and, for example, distortion or shape, may be important. These factors correspond to the problems and influencing factors evolving from the used TR collected in this work (see Table 2.5 in Section 2.3.2). However, the identification and the tighter definition of the extent of influences require further exploration for both: the perceived QF on 3DVO quality and the (subjective) quality assessment methods investigating them (e.g., quality assessment without reference material). Taking this and the problems with and influences on quality listed in Table 2.5 (see Section 2.3.2) (i.e., contortions, aliasing, blurriness, holes, superpositions, and disocclusions) into account, the following exemplary QFs in the context 3DVO can be defined, to date, as:

aliasing
blurriness
contortions
disocclusions
(un)naturalness (see section 3.2.2)
holes
shape deformations
superpositions
distortions

Video Communication-related Characteristics

Goldstein [101] describes the importance of cognition for social conversations. He mentions that this functional area of cognition includes all tasks that are necessary to interact with another person. These comprise the recognition of single persons, the perceptual processes in talking and listening, together with nonverbal communication through facial expression, gesticulation, and movement.

Argyle and Cook [71] report that, in a conversation, the amount of blinking or facial muscle activity strongly influences the perceptions of the communication partners and their state, in addition to utterances and head movements, mutual gaze or eye contact, with their recognizable information, for example, about pupil dilation, eye expression, or the direction of gaze-breaking. Accordingly, eye contact accounts for 30% of the overall gaze activity in a one-to-one conversation and is seen as the indicator that establishes relationships between people.

Rogers et al. [116] discuss technology-mediated social phenomena, present different kinds of computer-mediated communication systems, and investigate core aspects of sociability, communication, and collaboration by examining the main social mechanisms involved in modern media usage. These can be seen in relation to the use case eye contact supported video communication systems for private and business-related conversations, introduced in Section 2.3.1. Thus, experiences and emotions play an influencing role, as investigated in communication theories focusing on the notion of sociability (see, e.g., Simmel and Hughes [117]). Aaltonen et al. [118] reduce the mediated communication experience to five dimensions. These are: (1) emotional involvement, which includes factors such as enjoyment, importance, interest, valence, attention, or playfulness, (2) active participation, containing factors such as interactivity, control, arousal, perceived emotional interdependence, (3) reciprocity with the factors perceived behavioral interdependence, perceived affective understanding, and attention allocation, (4) co-presence, of which naturalness, personal, realness, and spatial relationships are factors, and (5) group cohesion. The latter one includes factors such as perceived message understanding, self-categorization, attraction to in-group, or commitment to in-group.

Nevertheless, and especially in connection with newly developed communication systems, the reason why something is rated as good or bad is often unknown and the definition of which factors are important and to what extent further research is needed. To reach the definition of certain video communication related characteristics influencing the quality also requires attention. This is also supported by the Research Group Entertainment Computing at the University of Vienna, Austria, as published by the Dagstuhl seminar 12181 in June 2012 [119] and recently by Pitrey and Hlavacs [120]. Based on the experience gathered with the TR of 3DVO created for the use case video communication and eye contact support (presented in Section 2.3.2), time delays may play a major role in the interpretation of facial expressions related to the conversation.

To summarize, based on influences from video communication applications, in this context relevant QFs could be, for example:

arousal
attention and attention allocation
attraction
commitment
control
enjoyment
eye contact
importance
interactivity
interest
naturalness
perceived affective understanding
perceived emotional interdependence
perceived behavioral interdependence
perceived message understanding
personality
playfulness
realness
self-categorization
spatial relations
time delay
valence

3.2.3 Summary of Influencing Characteristics and QF

The eye, its physical principles, and the related perception and cognition form the basis for the definition of the Human Visual System (HVS), which needs to be considered within quality assessment activities. Based on the TR of the regarded system and its usage context, various influencing characteristics are more or less relevant. At this stage of research, these characteristics have to be detected and carefully defined, taking into account the phenomena and humans' way of judgment, including all experiences and expectations. This subsection on QFs presents characteristics derived from perception and cognition, the use case 3DVO, and the usage context video communication.

Taking into account the general remarks of quality judgment, as described at the beginning of this section, and influencing characteristics from the use case related to visual perception (see Section 3.2.2), or, more precisely, to 3DVO (see Section 3.2.2), and herein to the usage of 3DVO in video communication (see Section 3.2.2) leads to the conclusion that the definition of QFs can be complex and always depends on the use case. Hence, possibly important QFs collected to date in this context are summarized in the following:

aliasing
arousal
attention and attention allocation
attraction
blurriness
contortions
control
commitment
disocclusions
distortions
enjoyment
eye contact
holes
importance
interactivity
interest
naturalness
perceived affective understanding
perceived emotional interdependence
perceived behavioral interdependence
perceived message understanding
personality
playfulness
realness
self-categorization
shape deformations
spatial relations
superpositions
time delay
valence

The overall quality of 3DVO is defined as an entity based on QE and QF, their relationships, and the system's scalability. The QEs are defined by the system's functionality; the QFs are defined by the users' judgment that is affected by their own quality perception. Influences on quality, based on usage context and environmental issues, have to be considered and assigned, accordingly, to the perceptual domain (e.g., pre-experiences by the user) or the physical domain (e.g., room changes). Taking into account the knowledge about QF and the ability to predict relationships between QE and QF, there will be a benefit e.g., for applications asking for scalability and adaptive TR, as more detailed information about relationships can be used to follow better applicable regulations.

3.3 Quality Evaluation

As outlined in Section 1.2.2, the current work aims at the evaluation of QoE and, especially, at the definition of influencing QFs in the use case 3DVO in video communication. Here, the question deals with choosing the evaluation design for the application, which is presented in Chapter 4. The following sections provide a review of existing evaluation methods for modern visual representations, considering (1) metric-based (see Section 3.3.2) and (2) subjective assessment-based (see Section 3.3.3) evaluation.

In the first section, Section (3.3.1), basic notions of statistics and the description of various experimental designs that investigate visual representations play a role. This includes the definition of evaluation, the description of common methods for both the investigation of overall quality and QoE including QFs, and the presentation of actual trends (see also Section 1.2.2). Some of presented methods may also be applied to focus on QoS. However, this is a matter of research design and the underlying research question. Based on this, methods which are judged as being useful for the definition of QFs are applied in Chapter 4.

3.3.1 General preliminary Remarks and Definition of Terms

Evaluation is defined as “objective and technically correct assessment” in the German Duden dictionary [121]. The word itself is derived from French (évaluation) and means *estimation.*

Quality Evaluation measures the quality of a certain product, service, or system. Taking into account the definition of *quality* given by Jekosch [5] (see also Section 3.1), quality evaluation is useful in order to assess what influences quality, and, consequently, to define the overall quality.

In order to assess such influences, several best practices are commonly employed; they can be differentiated according to whether they make use of:

- objective measures or subjective assessments or mixed method approaches
- quantitative or qualitative data
- categorical or measured variables
- tests with and without reference material
- test stimuli with low, high, or specified quality ranges
- laboratory tests and/or tests based on different contexts and usage environments
- direct versus paired comparison scaling or different ways of scaling

- univariate or multivariate data analysis
- parametric or non-paramtetric tests
- related or non-related data

The three methods of evaluation mentioned above (*objective measures, subjective assessments, mixed methods*) are outlined in more detail in the next sections. The bullet points listed above are relevant for the choice of the methodological application in this work. Therefore, they are described and discussed in the following.

Data left in their original form e.g., speech, text from observations or interviews and without quantification are defined as qualitative (QUAL) and data in numerical form, i.e., the results of measurement, are quantitative (QUAN) (see Coolican [99], p. 26).

Discussion of the measurement of variables leads to the question about which kinds of variables are differentiated. There are *categorical variables* providing information about a certain class, e.g., whether people use an Liquid Crystal Display (LCD) shutter-system, or an LCD polfilter-system, or a plasma shutter-system. In contrast, *measured variables* use a number to indicate on a scale where the answers of a test person are located on the scale for this variable. This applies, for instance, for a stereo test, in which people are tested for their ability to see stereo visualization, or for a test person tested for the degree of quality between the extremes good and bad. However, as Coolican [99] reports:

> "[s]ome variables are familiar in concept but measuring them numerically seems a very difficult, strange or impossible thing to do, as in the case of extroversion, attitude, anxiety or feelings. [...] Of course you can't *just* put a number on feelings - their qualitative differences are their most important aspect - but you can in some sense talk about *degree*. [...] how exactly we can measure, or at least somehow assess, such psychological variables as *attitude* and *anxiety*. We can start by asking, just what are these variables?"

Based on this statement, hypothetical constructs, as references to assumed inner states, may be involved in all the effects on human behavior, e.g., the amount of lip-biting, measured heart rate, or other references.

Whether tests are conducted *with or without reference material* usually depends on the availability of adequate reference material. Nonetheless, if reference material with which the test material is compared is not provided to the test participants, there is no guarantee that a test participant has an internal reference in mind. Taking into account the theory of visual intelligence, as described by Hoffman [122], makes the definition of *internal reference* even more complicated.

In practice, there is usually no choice between *low or high quality range* testing, because it is based on available stimuli and the development which should be tested. However, there are several references available that should be considered, in order to define to which quality range the stimuli tested belong and how to cope with these conditions in the evaluation design and statistics (e.g., Watson and Sasse [123], Winkler and Faller [124] and [125], Knoche et al. [126], Zhai et al. [127]).

This, and other aspects of the whole evaluation process as well, might be influenced by the evaluation environment (whether it is tested in a standardized laboratory or in the field) (see Jumisko-Pyykkö [95]), by the *context of use and the usage environment*, and, especially, by the task itself; e.g., when asking for active interaction (see Ninassi et al. [128]).

Taking all the above-mentioned aspects into account would influence the *way of how scaling is performed*: direct vs. paired comparison scaling, or single stimulus scaling (reference-free procedure) vs. double stimulus scaling in which test and reference image are displayed sequentially and both rated on two separate scales (1–10). Thus, with the same contextual effect, or comparison scaling when test and reference image are displayed sequentially, and ask for a difference rating on single scale (-10 to +10), with no contextual effect (see de Ridder [109]). These considerations about the number of categories, the number of stimuli, and the method of presentation are necessary, in order to avoid category effects with rating scales (see Parducci and Wedell [129]).

Following data collection, the correct way to analyze the data matters. The *univariate analysis* of data pays attention to the individual variable of a given data set, whereas the *multivariate analysis* considers each variable to be one dimension within a multidimensional space and allows for further analysis of the relationships between them (see Naes and Risvik [130], p. 21).

Parametric tests within the data analysis help to draw more conclusions, because the shape can be described mathematically if the data have an underlying (normal) distribution and an independent data set relationship. If this is not the case, the data are non-parametric and *non-parametric tests* are applied (see Coolican [99], p. 396). Similar rules for data treatment exist for the differentiation between *related* and *non-related* data (see Coolican [99], p. 79; see also Section 3.3.1).

Hence, in order to make decisions about all of these issues, the choice of the proper approach to the entire measurement process is based on the underlying hypothesis. Depending on the research question and the methodological design, the decision about the type of test participants is made i.e., expert viewers, naïve participants, consumers, semi-professionals, early adopters, etc., (see also Speranza et al. [131], or Coolican [99]).

The various measurement methods can be categorized into objective measures or subjective assessments, or mixed method approaches, as described in the following:

3.3.2 Objective Measures

Quality evaluation of visual representations is traditionally conducted via quality metrics (objective measures), which are either pixel-based metrics or based on models of human vision.

Very common pixel-based methods are MSE, which pays attention to image differences, and PSNR, which focuses on image fidelity, based on a reference image. They were developed especially for the consideration of luminance information. These approaches compare images pixel by pixel and the computation of these metrics is easy and fast (see also Equation 3.2 for MSE and Equation 3.3 for PSNR), but the relationship to the human quality perception is very limited (see Winkler [11]). Winkler describes MSE and PSNR as follows:

$$\text{“}MSE = \frac{1}{TXY}\sum_{t}\sum_{x}\sum_{y}\left[I\left(t,x,y\right) - \tilde{I}\left(t,x,y\right)\right]^{2} \tag{3.2}$$

> The MSE computes the mean of squared differences between gray-level values of pixels in two pictures or sequences I and $\tilde{I}$ for pictures with the size $X \times Y$ and T frames in the sequence. The PSNR is defined in dB as:

$$PSNR = 10\log\frac{m^{2}}{MSE}, \tag{3.3}$$

> where m is the maximum value a pixel is able to take.” [11]

In addition to the description of pixel-based metrics, Winkler [11] also provides conclusions why such error measures are not accurate enough for quality evaluations of modern visual representations: They are not able to reliably predict quality perception across different scene and distortion types.

Well-known quality metrics based on models of the human vision are either

- *(1) Relatively simple and computationally efficient single-channel models*: They consider, for example, spatial, temporal, and chromatic models of human contrast sensitivity in a defined color space e.g., CIELAB standard, as described in Wyszecki and Stiles [132].
- *(2) Multi-channel Models* such as the Visual Differences Predictor (VDM) according to Daly [133], the Sarnoff Just Noticeable Difference (JND) described by Lubin and Fibush [134], the Moving Picture Quality Metric (MPQM) introduced by van den Branden Lambrecht [135], the Normalization Video Fidelity Metric (NVFM) presented by Lindh and van den Branden Lambrecht [136], or the metric for Continuous Video Quality Evaluation (CVQE) of Masry and Hemami [137]. These are more general but accurate models.
- *(3)*, *Specialized Metrics* considering knowledge about used compression algorithms and special artifacts, as well as specialized vision models for a given application field such as the Digital Video Quality (DVQ) metric presented by Watson et al. [138], or the Video Quality Metric (VQM), as outlined by Wolf and Pinson [139].

Most of these metrics still require a reference as input and no metric is able to consider different purposes and replace subjective testing. These and several more metrics are presented and discussed in more detail in [11]. They mainly focus on the application for image compression, image quality, and video quality. In his work on the quality of digital video, Winkler [11] points out that “quality as it is perceived by a panel of human observers (i.e., Mean Opinion Score (MOS)) is the benchmark for any visual quality metric”. Moreover, he provides a considerable collection and discussion of metrics describing quality and vision models. He presents a collection of publications on the application of vision science to image processing via single-channel and multi-channel models (see Table 3.1):

Table 3.1: Overview of visual quality metrics after Winkler [11], p. 60.

Reference	Appl.[1]	Color Space [2]	Lightness[3]	Transform[4]	Local Contrast	CSF[5]	Masking[6]	Pooling[7]	Eval.[8]	Comments
Mannos and Sakrison (1974)	IQ, IC	L	$L^{0.33}$			F		L_2	R	
Faugeras (1979)	IQ, IC	AC_1, C_2	log L			F		L_2	E	
Lukas and Budrikis (1982)	VQ	L			yes	F	C	L_p	R	
Girod (1989)	VQ	L			yes	F	C	L_2, L_α		Integral spatio-temporal model
Malo et al. (1997)	IQ	L	?			F		L_2	R	DCT-based error weighting
Zhang and Wandell (1996)	IQ	Opp.	$L^{1/3}$	Fourier		F			E	Spatial CIELAB extension
Tong et al. (1999)	VQ	Opp.	$L^{1/3}$	Fourier		F		L_1	R	Spatio-temporal CIELAB extension
Daly (1993)	IQ	L	yes	mod. Cortex		F	C	PS	E	Visible Differences Predictor
Bradley (1999)	IQ	L		DWT ($DB9/7$)		W	C	PS	E	Wavelet version of Daly (1993)
Lubin (1995)	IQ	L		2DoG	yes	F,W	C	$L_{2,4}$	R	
Bolin and Meyer (1999)	IQ	Opp.		DWT (Haar)	yes	?	C	$L_{2,4}$	E	Simplified version of Lubin (1995)
Lubin and Fibush (1997)	VQ	L^* u^* v^*	yes	2DoG	yes	W	C(?)	L_p, H	R	Sarnoff JND (VQEG)
Lai and Kuo (2000)	IQ	L		DWT (Haar)	yes	W	$C(f, \varphi)$	L_2		Wavelet-based metric
Teo and Heeger (1994a)	IQ	L		steerable pyr.			$C(\varphi)$	L_2	E	Contrast gain control model
Lindh and van den Branden Lambrecht (1996)	VQ	L		steerable pyr.		W	$C(\varphi)$	L_4	E	Video extension of above IQ metric
van den Branden Lambrecht (1996a)	VQ	Opp.		mod. Gabor		W	C	L_2	E	Color MPQM
D'Zmura et al. (1998)	IQ	AC_1 C2	?	Gabor	?	W	C(?)		E	Color contrast gain control
Winkler (1998)	IQ	Opp.		steerable pyr.		W	$C(\varphi)$	L_2	R	See Winkler [11] sections 4.2 and 5.1
Winkler (1999b)	VQ	Opp.		steerable pyr.		W	$C(\varphi)$	L_2, L_4	R	See Winkler [11] sections 4.2 and 5.2 (VQEG)
Winkler (2000)	VQ	various		steerable pyr.		W	$C(\varphi)$	various	R	See Winkler [11] section 5.3
Masry and Hemami (2004)	VQ	L.		steerable pyr.		W	$C(\varphi)$	L_5, L_1	R	Low bitrate video, SSCQE data
Watson (1997)	IC	YC_B C_R	L^γ	DCT		not specified	C	L_2		DCTune
Watson (1998), Watson et al. (1999)	VQ	YOZ		DCT	yes	W	C	$L_?$	R	DVQ metric (VQEG)
Wolf and Pinson (1999)	VQ	L					Texture	H, L_1	R	Spatio-temporal blocks, 2 features
Tan et al. (1998)	VQ	L				F	Edge	L_2	R	Cognitive emulator

---

[1] IC: Image Compression, IQ: Image Quality, VQ:Video Quality.

[2] L: Luminance, Opp.: Opponent colors.

[3] γ: Monitor gamma, L: Luminance.

[4] 2DoG: 2nd derivative of Gaussian, DB: Daubechies wavelet, DCT: Discrete Cosine Transform, DWT: Discrete Wavelet Transform, WHT: Walsh-Hadamard Transform.

[5] F: CSF filtering, W: CSF weighting.

[6] C: contrast masking, C(f): ... over frequencies, C(φ): ... over orientations.

[7] H: Histogram, L_p: $L_p - norm$, exponent p, P_s: Probability summation.

[8] E: Examples, R: Subjective ratings.

These models and metrics are described detailed in Winkler [11], pp. 56–64, and used here only as summary of a collection of available approaches. In addition, subjective ratings, which are described in the next section, are also considered. Useful for the current work is the variety of characteristics of these metrics and the different metric components, which will have to be taken into consideration when it comes to the proposed interlinking model refinement in Section 5.2.9. Perceptually demanding visualization requires the consideration of several more factors.

The Video Quality Expert Group (VQEG) attempts to validate objective models of video quality assessment and therefore regularly conducts several tests and evaluations with subjective methods for validation (see VQEG Report [140]). Publicly available and most recent activities presented are in the context of television.

With respect to the objective measurement of stereoscopic television images, Kayser [112] provides a principle using weighting factors that have to be deduced from subjective assessment once, and are then summarized in order to rate the visualization quality objectively. A correlation factor of approximately 0.80 between the subjective assessment and the objective quality measurement, accounting for the detected weighting factors, demonstrates the reliability. Kayser [112] defines the quality of data compressed stereoscopic 3DTV images as a combination of psychovisual and physical parameters, leading to a certain impression based on the user-relevant aspects. In order to define a quality metric, he elaborated numerous parameters and conducted a comprehensive factor analysis. Rittermann [26] presents this procedure in four steps. Firstly, subjective assessment aspects are defined with the usage of a predefined questionnaire based on a representative image data base. Secondly, the dimensions of the subjective perception are determined by using a factor and cluster analysis. In a third step, a precisely traceable mathematical model that considers all relevant aspects is created. Finally, a quality measure is defined from the results of the factor analysis, based on the mathematical models. Subsequently, the quality aspects elaborated in this way are reduced to the modeled aspects. For these modeled aspects, several models for their estimation are developed. This occurs with different approaches (e.g., frequency analysis, variance analysis, block estimation). The merging into a quality measure occurs with consideration of weighting factors that have been detected by the factor analysis. (See also Rittermann [26] p.56.) This approach is included within Ritterman's 3DVO quality definition. Rittermann [26] presented a 3D Video Object Quality Metric (3DVQM) in order to define the quality of a 3DVO. This mathematical measure includes quality influencing factors, which are defined by objective measurements, to a large extent, and verified by subjective assessment, to a small extent (see Section 3.3).

According to [26] the 3DVQM uses calibrated views for the computation of statistical parameters and the detection of distortions. These parameters are combined and lead to the 3DVQM (see Equation 3.4). The 3DVQM uses a weighted sum of quality parameters (p_ψ).

$$3DVQM \equiv \sum_{i=1}^{N} \beta_i \cdot p_{\psi,i} \tag{3.4}$$

Rittermann [26] points out that this metric's accuracy depends on the coefficients (β_i) chosen. Therefore, a change involving the quality parameters ($p_{\psi,i}$)(e.g., they are either modified or enhanced) also leads to a change of the related coefficients as weighting or limiting factors. This fact shows the openness of this metric for innovations concerning the 3DVO TR as well as its quality judgment.

Within the thesis, it is intended to extend this 3DVQM and duly appropriate it by the inclusion of extensive subjective assessments of quality. This involves the collection of more information about the quality influencing factors from a perceptual point of view including information about why something is rated as good or bad. As previously described in Section 2.3.2, the TR is still in progress (i.e., there is no generalizable variety of test data available) and several relevant aspects are present interdependently. Therefore, the definition of 3DVO quality still has to be approached by gradual approximation, as also suggested by Rittermann [26]. Once relevant aspects are detected, a weighting has to be performed, in order to grade the influence (leading to coefficients, β_i). Rittermann [26] points out that this weighting has to be created by an objective evaluation and verified by subjective tests oriented to the ITU-T recommendation on Subjective Video Quality Assessment Methods for Multimedia Applications (ITU-T P.910) [141]. Hence, quality parameters are predefined and measured with weighting factors. The presented measure mainly includes quality influencing factors under usage of statistical units (parameters, $p_{\psi,i}$) derived from the TR, including *wrong angles*, *silhouette errors*, *synthesis errors*, local vs. global errors that are based on, e.g., *occlusion*, and the *dynamics* of these quality factors.

Here, initially examined subjective quality factors, such as object areas that seem to be defective, contour errors, position faults/commotion, and a general blur of the object, emerge, for example, from technical factors such as occlusion, distortion, and shape. However, in Rittermann [26] it is also outlined that the context of use, special usage tasks, the user's expectations, and the definition of quality influencing factors (defined termini or attributes describing quality) have to be considered. Therefore, the identification and the tighter definition of the extent of influences, as well as extensions caused by further technological developments, demand further exploration and adequate assessment methods.

Solh and AlRegib [142] present a novel Multi-view Image Quality Measure (MIQM). They introduce this measure as a combination of three index measures that quantify the physical nature of multi-view image distortions. These three index measures, namely (a) the luminance and contrast index, (b) the spatial motion index, and (c) the edge-based structural index, are multiplied with values ranging from 0 for maximum distortion to 1 for minimum distortion. This measure was validated using single-view images. However, in order to validate it against, e.g., MOS of multi-view images, adequate data were still lacking.

Gastaldo et al. [143] describe activities aimed at the objective quality assessment of displayed images, using neural networks to mimic perceived image quality. They state the necessity for objective methods to match the perceived image quality as measured by subjective testing. Subjective assessment is described in the following section.

3.3.3 Subjective Assessment

Winkler [11] points out that "[s]ubjective experiments represent the benchmark for vision models in general and quality metrics in particular". The traditional measure of use in order to validate the performance of objective metrics in visual quality evaluation is the MOS (see also ITU recommendations on subjective assessment of quality (ITU-T BT.500-13) [36]). Several ITU standards that focus on subjective assessment methodologies are available for different purposes. Distinctions are made between the method of stimulus presentation, the method of assessment, and whether there is an explicit, a hidden, or no reference presented to the test participants (see method overview in Table 3.2). These are common methods using subjective assessment in the area of visual quality evaluation.

In 2009, Winkler [144] discusses the distribution and variability of subjective ratings in video quality experiments, the effects of discrete rating scales, and the number of subjects needed.

Pinson and Wolf [145] compare several subjective video quality testing methodologies. They consider data from six different subjective video quality experiments that were performed with the Single Stimulus Continuous Quality Scale (SSCQE), the Double Stimulus Continuous Quality Scale (DSCQS) and the Double Stimulus Impairment Scale (DSIS) methodologies (for examples see Table 3.2).

The methodologies were differentiated and compared with regard to their described advantages and disadvantages, especially in terms of the length of presentation of the used stimuli. Generally speaking, they found no large difference in the results of each method, but an influence of stimuli length was observed. Pinson and Wolf [145] point out that they "[. . .] provide evidence that human memory effects for time-varying quality estimation seem to be limited to about 15 seconds." Against this background, it is suggested that results based on different forms of stimuli presentation and assessment scales are comparable with regard to the length of rated sequences and the relevant part of forming an estimation.

Table 3.2: Methods of Subjective Quality Evaluation, after Kozamernik et al. [146].

Parameter	SS[1]	DSIS[2]	DSCQS[3]	SSCQE[4]	SDSCE[5]	SAMVIQ[6]
ITU-R Standard	BT.500-11	BT.500-11	BT.500-11	BT.500-11	BT.500-11	BT.700-11
Explicit reference	no	yes	no[7]	no	yes	yes (uncompressed)
Hidden reference	no	no	yes[8]	no	no	yes
High anchor	no	no	yes[9]	no	no	no (hidden ref)
Low anchor	no	no	yes[10]	no	no	yes
Scale	5 grades	5 grades	bad - excellent (continuous quality scale)	bad - excellent (continuous quality scale)	bad - excellent (continuous quality scale)	bad - excellent (continuous quality scale)
Sequence length	10 s	10 s	10 s	=5 min	10 s	10 s
Picture format	all	all	all	all	all	all
2 simultaneous stimulus	no	no	no	no	yes	no
Presentation of test material	once	variant I: once variant II: twice in succession	twice in succession (double stimulus)	once	once	as often as user likes (multi stimuli)
Videos per trial	1	2	2	1	2	max 10[11]
Voting	only test sequence	only test sequence	only test sequence and reference	only test sequence	difference between the test sequence and ref. simultaneously shown	test sequence and reference
Possibility to change the vote before proceeding	no	no	no	no	no	yes
Continuous quality evaluation	no	no	no	yes (moving slider in a continuous way)	yes (moving slider in a continuous way)	no
Minimum accepted votes	15	15	15	15	15	15
Rejection criteria	o	o	yes, but not stable	o	o	yes
Observers per display	1 to many	1 to many	1 to many	1 to many	1 to many	1
Display	all (mainly TV)	all (mainly TV)	all (mainly TV, DLP)	all (mainly TV)	all (mainly TV)	all (mainly PC, PDA)
Quality results	relative	relative, depending on reference quality	relative, depending on compared sequence	relative	relative, depending on ref quality	absolute measure of video quality

---

[1] Single Stimulus (alike Absolute Category Rating (ACR) in ITU-T.P.910).

[2] Double Stimulus Impairment Scale (alike Degradation Category Rating (DCR) in ITU-T.P.910).

[3] Double Stimulus Continuous Quality Scale.

[4] Single Stimulus Continuous Quality Evaluation.

[5] Simultaneous Double Stimulus for Continuous Evaluation.

[6] Subjective Assessment Methodology for VIdeo Quality.

[7] Is not mandatory (could be any test sequence).

[8] Is not mandatory (could be any test sequence).

[9] Is not mandatory (could be any test sequence).

[10] Is not mandatory (could be any test sequence).

[11] Different bit rates in one trial (to avoid contextual effects).

Besides these standardized methods, several additional approaches investigating subjective assessment of video quality are available. For example, Pereira [147] includes sensations and emotions and consumer electronics video adaptations. Radun et al. [148] focus on the explanation of multivariate image quality and an interpretation-based quality approach. Reiter and Köhler [149] investigate the subjective assessment of bimodal perception in interactive audiovisual application systems, and Rouse et al. [150] envisage tradeoffs in subjective testing methods for image and video quality assessment. Qualitative descriptive quality evaluation methods for naïve participants are introduced by Lorho, using an individual vocabulary profiling approach (see also Lorho [151], or Lorho [152]).

Furthermore, in addition to comparing data collected in different subjective assessment procedures, comparisons are also useful with objective measures or with data derived from evaluations that collect quantitative or/and qualitative information (see also Borrego et al. [153]). Methods for defining thresholds can also be part of mixed method approaches, for instance, the Alternative Forced Choice (also called Adaptive Forced Choice or IFC - Interval Forced Choice) (AFC), or the method of limits approaches such as, for example, used in psychoacoustics and described by Levitt [154], or as suggested by Fechner [155] for visual use cases (see Knoche et al. [126]). Several common approaches in the field of multimedia quality evaluation are described in the next section that deals with mixed method designs.

3.3.4 Mixed Method Designs

Mixed method designs are useful in order to obtain additional information on QUAN and QUAL data and, moreover, to relate or compare this information in order to reach a broader perspective on a certain research topic. Creswell and Plano Clark [156] categorize different mixed method designs based on their design, design pattern, and general purpose, as outlined in Table 3.3. Free Choice Profiling (FCP) methods and Lorho's descriptive approach that uses individual vocabulary profiling include naïves (see Lorho [151]). This table provides an overview of methodological designs and for which evaluation purpose they are useful. Furthermore, it informs about which kind of data collection is adequate. One example for a mixed method design is the Open Profiling of Quality (OPQ) method that includes ratings from a quantitative (psychoperceptual) evaluation of hedonic excellence and naïve participants' individual vocabulary which is

Table 3.3: Mixed method designs, according to Creswell and Plano Clark, according to Strohmeier [157].

Mixed method design	Design pattern	Purpose
Triangulation design	Independent collection of QUAN and QUAL data. Interpretation based on both data sets.	Comparison of QUAN and QUAL results for a broad interpretation of the results
Embedded design	One data set is used in a supplemental role in studies primarily based on the other data set.	Additional qualitative expressions about quantitative results (e.g., supporting decisions about further studies or tasks)
Explanatory design	Two-step design. First collection of QUAN, then QUAL.	QUAL data may be needed to explain unexpected results or to detect errors in the QUAN research design.
Exploratory design	Two-step design. First collection of QUAL, then QUAN.	QUAL data may be needed to explain unexpected results or to detect errors in the QUAN research design.

used for a descriptive sensory analysis (see Strohmeier [104]). Another descriptive approach is the RaPID perceptual image description method (RaPID) that was already introduced in 1996 by Bech et al. [158]. Here, vocabulary for describing quality is defined by experts as test participants.

Mixed method approaches are various method combinations within an experiment. They may constitute, for example, interviews such as expert interviews, online questionnaires (see Coolican [99]), or focus groups (see Morgan [159]). Focus groups are methods in qualitative research that also involve group interviews to discuss, include personal experience, and comment on the topic of research. To explore a research topic, these methods seem to be useful, too, just like literature analysis (e.g., Robinson [160], Glaser and Strauss [161]). The choice depends on gathering the data that might be most relevant for the research. Therefore, the correct choice of analysis is mandatory (see Chapter 4).

3.3.5 Sensory Evaluation / Multivariate Analysis

Sensory evaluation and its respective multivariate analysis methods are presented here in brief, because they will play a role in the following chapter. These focus on the evaluation with descriptive methods that collect perceptual attributes of the perceived quality.

For the description of the OPQ method, Strohmeier [104] provides a framework for the analysis of sensory data using multivariate analysis. Lorho [151] suggests analyzing perceptual attributes in several steps: (1) presentation of individual attributes collected by a general and very qualitative clustering and grouping, (2) attribute grouping by hierarchical cluster analysis, leading to a dendogram that indicates dissimilarity and similarity of individual attributes at different levels, (3) analysis of perceptual differences between algorithms using the Generalized Procrustes Analysis (GPA), which leads to a group average configuration (matrix of average scores on a common set of underlying attributes) and information about samples, attributes, and test participants. Based on this, detailed information about the method of analysis chosen within this work is presented in the relevant sections about the laboratory experiment in Section 4.2.2, as well as the used OPQ, described in Section 4.2.3.

3.3.6 Assessment of Quality Influencing Factors

The goal is to detect QFs and their degree of influence. Therefore, after a definition of the influencing factors (see Section 3.2), it is necessary to establish how they are evaluated.

Factors in an Experiment

According to Coolican (see Coolican [99], p.77), independent variables are also known as factors. This definition becomes important when discussing more complex (e.g., multi-factorial) designs. Within this work, the following definitions are used:

QEs as explained in Section 3.2.1, are based on the TR and will not be evaluated in detail here (i.e., in terms of detecting them). Instead, these factors are defined by the technical approach used and are seen as independent variables. Experts (see Section 4.2.1) define whether a QE has a more or a less massive influence on the QoS.

The activities described subsequently will deal with questions concerning the definition of QFs, whether they are less dominant or factors exhibiting greater influence, and their relationship to the TR for the special use case (see Chapter 2.3.2).

Assessment of Quality Features (QF)

Perception-based factors may have different sources and reasons for experiencing them. As already described in Section 3.2.2, Strohmeier [104] points out that "[t]hese quality perceptions encompass both low-level sensorial and high-level cognitive processing, including knowledge, emotions, attitudes, and expectations". Here, the definition of less dominant versus factors with more impact depends not only on the HVS or pre-experiences. Considering this challenge, traditional methods for quality assessment are insufficient for measuring the QF of 3DVO for the use case eye contact support for video communication.

Available measures, for example, various threshold definitions mentioned above, are useful for the exclusive evaluation of common but special cases. They are useful for example to investigate the impact of particular coding methods as addressed in [162] and [29], or the method of transmission as demonstrated in [163], or the usage environment according to [95]). But they are not adequate for the special task of video communication and the assessment of the visualization quality of 3DVO to support eye contact. Case-based reasoning and its realization, the eye model related to form and shape recognition, human light sensitivity and, thus, the right way to monitor calibration, and many other reasons why something happens, have to be taken into account. However, the task and the use case may play the most important role (see Rogowitz and Goodman [13]).

3.3.7 Summary - Evaluation

For quality evaluation, several evaluation and data analysis methods are available. Based on the research question, one can choose between objective measures, quantitative (psychoperceptual), qualitative (descriptive), and mixed method approaches. This may be influenced by the task and/or the use case. For the assessment of QFs and the definition of the visualization quality of 3DVO to support eye contact in video communication, the described traditional methods might not be applicable without constraints. Thus, the actual challenge for QF assessment is twofold. On the one hand, there is the evaluation without (pre-produced) reference material other than the user's experiences from real face-to-face conversations. On the other hand, there is the exclusion of real-time interaction and audio through the early development phase and the lack of a possibility to realize it technically. This means that it is not possible to define thresholds with traditional methods. However, it is assumed that different methods are available to collect data and information leading to the tighter definition of QFs of 3DVO quality. Purposes and criteria are defined in Section 4.1 in order to support the right choice of evaluation methodology. The definition of these purposes and criteria supports the choice of the correct method on a meta level (as introduced by Nowak [164], p. 56., in *Evaluating evaluation measures*).

4 Towards the Definition of Quality Features

This section describes the choice of the evaluation method, the methodological approach used, and the evaluation steps, in particular, in order to define important QFs. Based on the research question and the goal to identify relevant quality influencing factors deriving from the users' subjective quality estimation, the following evaluation design is used. This is motivated by the question on how the technical realization process of 3DVOs creation can be interlinked with the definition of resulting quality and, thus, profit from subjective quality assessment. Considering supplementary research questions on how to integrate the user's subjective quality estimation into the definition of the overall quality and how to identify the most relevant factors influencing quality, the decision for adequate evaluation and data analysis methods is presented here. When considering the context of the use case video communication, additional theories on sociability support and human perception need to be investigated (see, e.g., Section 3.2.3). This requires usefulness of the expected results to support the development of an extended quality model showing influences on, and relationships to QFs and QoEs that support technical optimization.

To detect to what extent the resulting quality of different processing steps (i.e., predefined QEs) is acceptable and to examine factors influencing the experienced subjective quality (i.e., QFs) of represented 3DVOs this work needs adequate methods which cope with a large amount of degrees of freedom. This degrees of freedom includes the experience of eye contact, and, in particular, the measurement of possible influences by several characteristics (e.g., appeal and trustworthiness) of the shown conversational partner, and visual representation related influencing factors based on the literature, e.g., as examined by Rittermann [26], or the description (outlined in Table 2.5) of possible problems within certain TR processes used by Weigel [2]). Therefore, the methodological approach is as follows: it starts with an explorative approach that contains two literature analyses, two focus groups and an online interview, and is then followed by an extensive laboratory experiment and an OPQ study, which allows the collection of broad and concrete information on quality influencing QFs (other than eye contact) and their importance. Before these applied methods are described in more details in Section 4.2, a short discussion on the definition of useful methods is given in Section 4.1.

4.1 Definition of Useful Methods

Theoretically, several methodological approaches for the evaluation such as objective measures, subjective assessments or mixed methods (see Section 3.3) are available. This makes the decision for the right evaluation framework complex. Therefore, a list of criteria and purposes is presented in Table 4.1, in order to facilitate the finding of the evaluation design and to choose an adequate methodological approach. The rationale of defining the questions is based on the requirements that the current work addresses. Hence, they are in line with the goals outlined in Section 1.1.2. The question arises whether it is possible to define such purposes and criteria to a more universal extent. Nowak [164] discusses this issue in another context in an extra section called *Evaluating evaluation measure* and points out that "often, the question about the desired outcome of an evaluation measure is neither easy to define nor to prove". In addition, Novak cites Dupret and Piwowarski (2010) who tellingly describe:

> "[d]eciding which metric is best calls for a third 'meta' metric. Because various 'meta' metrics are likely to co-exist, a meta metric for the meta metrics is necessary, etc.[...] On one hand, a performance measure should relate to the user model and evaluation objective [...]. On the other hand, necessary and desired requirements on metrics can be objectively defined [...]." [164].

Roto et al. [35] provide a guideline for the right method choice based on the addressed circumstances, e.g., development stage, research question, availability of reference material and test participants. Kunze et al. [165] introduce an extensive comparison model to guide between-method comparisons considering more than simply purposes and assessment-related criteria (e.g., duration, costs). This approach was considered in activities related to new methods, as in [157], in order to choose an adequate methodological concept as well. In the usability research area, and focusing on task analysis

(TA), Embrey [166] defines questions in order to decide whether a particular method meets specific purposes and specific selection criteria or not. Embrey [166] does not describe how the definition of purposes and selection criteria was realized. It seems that he derived the respective advantages and disadvantages of each mentioned method from a prior description. The current work investigates this approach against a wider background, including various evaluation methods that have been introduced in Chapter 3. The investigation focuses on the initial research questions and defines purposes and selection criteria, as in [166].

When designing an evaluation in order to answer a particular research question, it is not always obvious which method to choose. To test scientific theories, various practices are available, for example the hypothetico-deductive method (see Coolican [99]). For a subjective multimedia quality assessment, different guidelines, e.g., from the European Broadcasting Union, as described by Alpert and Evain [167] and standards (e.g., ITU-T BT.500 [36]) provide best practices in order to test specific research questions under special conditions. Currently available standards focus on a particular context and a specified use case, paying attention to pre-defined quantifiable values. Although these methods deal with subjective quality assessment, the results do not usually yield a deeper understanding about the users' preferences, i.e., why something is rated as good or bad. Therefore, a more user-centered approach seems to be useful. Focusing on evaluation methods to measure the QoE, activities are undertaken in order to complement quantitative standardized measures with sensory profiling (see Strohmeier et al. [157]), including the users' individual opinion about quality. This approach facilitates the investigation of QFs.

Activities on the definition of UX were collected and they presented adequate evaluation methods (see Vermeeren et al. [168]), and information about why and when to use the respective method (e.g., see Roto et al. [35] or Kunze et al. [165]). Similar activities have been carried out for the creation and further development of the definition of QoE (see Le Callet et al. [32]).

In this context, and especially in this thesis, the use case is twofold in a complex manner: One challenge is the complexity of the TRs and the produced quality, which influences the perceived quality and does not provide reference material for evaluation purposes. The other challenge is the definition of adequate evaluation methods in order to detect and define the influencing QFs, the extent of their influence, their relations, and a weighting of their influence, and out of this forming the perceptual domain. Different levels of defining the influencing factors may require different evaluation methods. In a first step, factors (especially QFs) need to be detected. Afterwards, their level of influence on the overall quality can and should be evaluated. Furthermore, links between the TR factors and the QFs should be determined, as well as their weighting in the overall influence.

Similar to the procedure by Kunze et al. [165], and as presented in [166], the following questions are defined.

Table 4.1: Purpose and selection criteria that a quality evaluation method should meet.

Does the method meet the purpose/criteria to...
detect quality features?
identify **all** important factors?
evaluate the degree of dependencies?
describe the quality features?
define relationships (between TR quality factors and quality features)?
classify the detected features based on severity?
investigate thresholds?
provide a qualitative description of important factors?
investigate the quality of experience?
analyze the determining factors?
investigate relations between influencing factors?
describe quality features quantitatively?
be comprehensive?
provide deterministic factors?

They are based on the common methods used for, and based on, the general purpose, which should be considered when choosing the appropriate method. In the following, purposes and criteria that the method should meet are presented in order to define the perceptual domain and its QFs and links to the TR of visual representations. This is presented here exemplarily, because this table does not aim at completeness but rather serves as a tool for choosing adequate evaluation methods at this stage.

This table excludes factors such as the time needed or other economic topics (e.g., as considered in [165]). The methods considered with these questions in mind are based on an extensive literature analysis and theoretical information that is also partly introduced and referred to in Chapter 3. These descriptions do not include methods such as DSCQS and a Subjective Assessment Methodology for Video Quality (SAMVIQ), because “real” reference material is not considered. Furthermore, IBQ (e.g., as utilized by Radun et al.

[169]) is not considered, because, in addition to the missing reference material, a realtime application is not available. These methods were not perceived as being more useful than other methods mentioned in Chapter 3 (i.e., Method of Limits, or reference-based methods), in order to address the research question of the thesis.

Although it is intended to define relations between single QFs and QEs and weighting factors as well, methods of limits, according to [155], and reference-based methods, e.g., the ITU-T P.910 [141], are not considered here. The reasons for this are the available evaluation possibilities that are based on the circumstances provided by the used TR (see Section 2.3.2) within the project Skalalgo3d. There, (realtime) continuous change in quality, which would be useful for a method of limits, and reference material (i.e., original representation or seen as ideal result) are not available.

No method covers all context- and pre-defined purposes and criteria (see also Table 4.1) that the evaluation activity should investigate. Hence, in the next step, a mixed method approach is considered. After the implementation of one method or another, the information available may then lead to other decisions about the most adequate method.

4.2 Evaluation Design

A mixed method approach is considered by taking into account the evaluation purposes and selection criteria mentioned in Section 4.1. The reason for this is that there is no method covering all context and pre-defined purposes and criteria, as shown in Table 4.1. The first method used was an explorative approach as a preliminary step, followed by an experimental large-scale laboratory experiment. Based on this, a second laboratory experiment was used in order to reach a quantifiable result on perceived quality influencing factors. This second laboratory experiment uses OPQ of further developed test material. The results of the OPQ study are factors and QFs that may have been already mentioned within the open question task of the first laboratory evaluation, but were not quantified in this first step. This combination of different methods intends to support the integration of subjective quality estimation into the development process. The overall evaluation approach is visualized in Figure 4.1 and contains three studies. Namely, these are Study 1, which includes a literature analysis, interviews (expert interview, online questionnaire) and focus groups, and two further experiments (Study

2: laboratory evaluation, and Study 3: OPQ). These lead to a final expert evaluation, which is part of Chapter 5, in which the quality taxonomy refinement is discussed. With the quality taxonomy a contribution to the TR by evaluation results is intended.

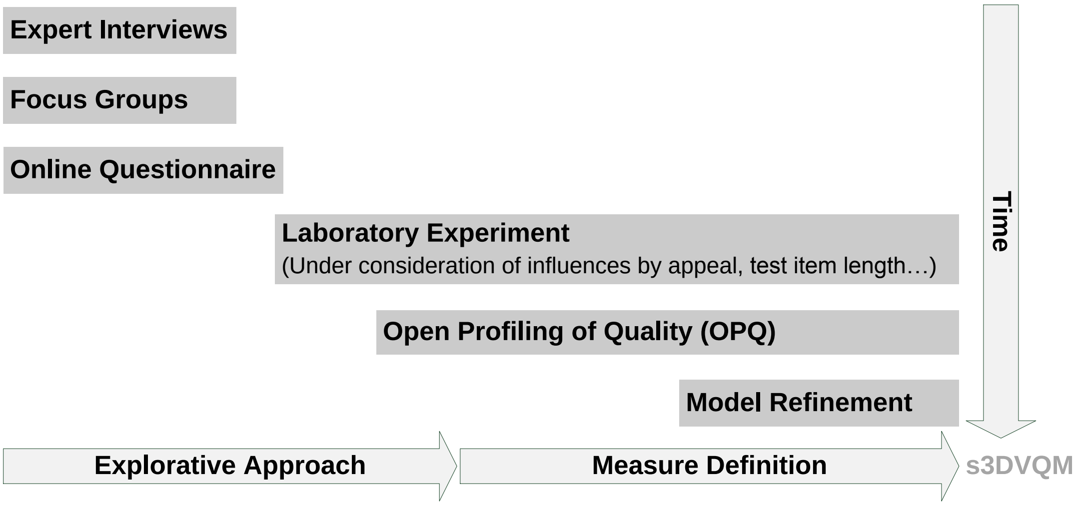


Figure 4.1: Evaluation Approach.

In the following, the depicted laboratory evaluations (Study 2 and Study 3) use the test items created by the different methods described in Section 2.3.2 and listed in Table 2.3 and 2.4. The five lines in bold type (S1, S2, and A1, A2, A3) describe the methods used for test items in Section 4.2.2 (the laboratory evaluation). The six lines highlighted in italics (S3, S4, S5, and A2, A4, A5) contain the methods used for the test items in Section 4.2.3 OPQ. A set of test items created within the project Skalalgo3d (see Section 2.3.2) is available. See description online at the project's blog at http://blogs.tu-ilmenau.de/skalalgo3d/.

4.2.1 Study 1: Explorative Approach

The aim was to obtain a broader understanding of the different understandings of quality, quality influencing factors, and the perceived quality in the area of 3DVO generation and consumption, and, moreover, information about existing approaches for interlinking QoE and TR. Towards this, the first study was devided into four steps: a literature analysis, three expert interviews, two focus groups with 15 participants in total, and a survey with an online questionnaire answered by 25 participants. Questions in the interviews, focus groups, and questionnaires addressed pre-experiences with 3DVO, the topic quality in conjunction with 3DVO creation and consumption, and ideas on possible future 3DVO applications in the area of video communication. Information about needs, wishes, and expectations for 3DVO, its quality, and possible further developments of the existing creation process was collected.

In the following, the methods used, the accompanying expected outcome, and their results are described in more detail.

Literature Analysis

There are different approaches for conducting a literature analysis (see Coolican [99], Robinson [160], or Glaser and Strauss [161]). In this work, two literature analyses, Literature Analysis A and Literature Analysis B have been conducted in two different ways, as described in the next paragraphs.

Literature Analysis A
In a first step, literature focusing on 3DVO creation and perception was collected and analyzed, with the intention of detecting the mentioned quality influencing factors (QF and QE). In doing so, an explorative, qualitative, and systematic approach was employed, collecting the mentioned QFs and QEs (using a grounded theory approach, as described by Glaser and Strauss [161]). This activity results in a conceptualization of quality influencing factors based on different TRs and works as a complementary activity to the processing and scene representation approaches described in Section 2.2 and their related problems and quality influencing factors.

Literature Analysis B
In the second step, a systematic literature analysis using the Survey, Question, Read, Recite and Review (SQ3R) method, based on Robinson [160], was conducted. This analysis was done in order to obtain a basic idea of the state of the art of interlinking QoE and TR, as well as different approaches to model and/or taxonomy development. This was conducted by first defining a number of search keywords. These keywords were based on preliminary general research into existing concepts to communicate evaluation results in the field of QoE (or similar findings). Subsequently, a list of the initially detected

concepts is presented. Terms (and alternatives) such as norms (standards, qualifications, rules), design patterns (guidelines, styleguides), recommendations, taxonomies, semiotics, and heuristics were used as search keywords for the substantial literature review. Additionally, terms derived from the usage context of quality evaluation of 3DVO representations for video communication were applied (e.g., requirements, designing experience, quality assessment).

Literature Analysis: Results

The following results are based on the two literature analyses focusing on 3DVO quality influencing factors (Literature Analysis A) and on the construction of an QoE and TR interlinking model (Literature Analysis B).

Results of Literature Analysis A
During the first literature analysis, literature focusing on 3DVO creation and perception (Ohm and Müller [56], Pollefeys et al. [27], Scharstein and Szeliski [60], Carranza et al. [61], Smolic and Kimata [48], Smolic et al. [29], Schreer et al. [38], Müller [59], Smolic et al. [44], Rittermann [26], Kühhirt and Rittermann [6], Grau et al. [66], Weigel et al. [10]) was analyzed with the aim of detecting the mentioned quality influencing factors (QFs and QEs); see also Table 4.2. With the aid of an explorative, qualitative, systematic approach, a list of quality influencing factors detected in the mentioned publications was created and the items were categorized using the grounded theory approach according to [161]. The results are presented in Table 4.2 and include both the source and the category of the collected factors.

Table 4.2: Quality influencing factors derived from literature.

Category: Quality influencing factor	Source(s)	Related TR process
Acquisition:		
lower quality based on inaccurate camera calibration (e.g., color differences, time and/or space dependent differences)	[66], [26]	camera setup and accompanying amount of cameras
equipment and complexity costs, increased processing	[29], [44], [26]	
2D quality influencing factors (time dependent)	[26]	
(Pre)Processing and Scene Representation:		Process related to source(s) and influencing factor
lack of natural illumination and reflectance changes if viewpoint is altered	[56], [27], [29]	Intermediate viewpoint interpolation, view-dependent texture mapping
computationally burdensome	[56]; [27], [61], [26]	3D modelling; viewpoint dependent texture synthesis
artifacts near separation position caused by unreliable disparity estimates, different illumination and reflection effects, or color deviations between the cameras	[56], [27], [29]	I3D: primary surface; *dynamic* texturing (opposite to static texturing)
slight distortions near the edges of the object	[56], [27]	I3D: synthesis of viewpoints
holes caused by false estimates producing discontinuities in the disparity information	[56], [27], [61], [26], [10]	I3D, voxel based representation
outlier/mismatches because of textureless regions, occluded regions, or depth discontinuity regions	[27]; [60], [61], [26], [10]	mapping; local/global stereo correspondence algorithms; stereo approaches; texture generation
poor segmentation leading to lower quality (background and foreground definition)	[61], [66], [10]	way of segmentation
wrong perspectives (view dependent failure, space dependent failure)	[26]	way of modelling
silhouette failures	[26]	way of segmentation
no realtime capability	[29], [26], [10]	disparity estimation algorithm
Coding:		
reconstruction errors	[61], [29]	kind of mesh coding (e.g., 3DMC, D3DMC, GOM11, AFX-IC (for details on this see Smolic et al. [29])
Interactive Rendering:		
poor rendering results	[29], [44]	static texturing

Table 4.2 mentions the related TR or 3DVO creation process leading to the listed factors only to a limited extent, even though these occur in the majority of the investigated publications. This broad categorization is based on the consulted literature sources and needs to be defined according to the TR actually used in the investigated use case. The categories may, thus, differ e.g., in the related project Skalalgo3d described in Section 2.3.2, the synthesis takes place before coding, transmission, and presentation, but the allocation between TR and quality influencing factor may be more accurate. However, the aim was to collect the quality influencing factors mentioned in the literature, in general, in a first step. Depending on the TR used in this work, a collection of problems that lead to a reduced or bad quality, related to the creation processes, is provided. This is presented in Table 2.5 in Section 2.3.2, based on Weigel [2]. A consideration of these listed factors and the defined categories, leads to questions about how these factors are perceived by the end user. A further question is whether the list is complete, whether the end user perceives more, or other, factors, and what would change if the TR changes.

Results of Literature Analysis B
During the second literature analysis, taking into account only keywords already mentioned (e.g., standards, requirements, taxonomies), 324,617 sources were found. The oldest source was published in 1883 and is about rules; the most recent publications are from 2011 and cover the entire range of mentioned keywords. Of these, and limiting the choice to the latest publications (published in the years from 2009 to 2011), as well as using the IEEE selection tool, 25 apparently (in terms of content) useful sources were selected for further analysis (i.e., Silzle [4], Hsueh et al. [170], Möller et al. [171], Ramanan and Baker [172], Chang, Chang, Chen, and Lei [173], Anh and Mellouk [174], Mangtang and Kecheng [175], Chaoqun [176], Bento et al. [177], Mittal et al. [178], Jong-Seok et al. [179], Moorthy et al. [180], Cerra et al. [181], Cardeal et al. [182], Du et al. [183], Gershon [184], Hyun-Jong and Seong-Gon [185], Wang et al. [186], Hyun-Jong and Seong-Gon [187], Lopes Gomes et al. [188], Yebin et al. [189], Ekmekcioglu et al. [190], Bruls et al. [191], Douglas et al. [192], ITU-T BT.500-13 [36], Cheng et al. [193]; these are the references [7]-[31] in Kepplinger and Weigel [18]). In summary, the findings published in [18] reveal that every investigated concept meets special functions and use cases and may be useful for defining relationships and dependencies between QoE influencing factors and their origins. However, for the purpose of identifying and breaking down various quality influencing features, their traceability and knowledge about their relationships (including a weighting) are necessary. Therefore, the concept taxonomy seems to be useful e.g., in combination with ontology, in order to overcome the disad-

vantage of calculability. Detailed results are presented in [18] and used in this work during the development of a model interlinking QoE and TR in Chapter 5.

Expert Interviews

Three experts working in the field of 3D technology and free viewpoint video production were interviewed. The two researchers and one film producer had similar pre-experiences. Their reason to become involved with the topic 3D video (objects) is motivated by professional engagement and private interests. Their work is basically related to coding and compression methods. Quality is important for all three experts. One of them is familiar with the European 3D Media Cluster (see, e.g., www.20203dmedia.eu, last viewed 4th Dec. 2013) and the Motion Picture Experts Group (see, e.g., Smolic et al. [44] or mpeg.chiariglione.org, last viewed 4th Dec. 2013). One has developed several 3D systems in order to produce and display 3D content taking into account subjective perception. The third expert is very familiar with stereo analysis, view synthesis, and 3D video communication. All experts received the same questions and were interviewed individually. The interview questions are presented in Appendix A. Questions in these interviews addressed their points of view on the merging of subjective quality factors with objective quality measurement, possible ways to do so, and the most important quality factors in the area of 3D and free viewpoint videos. Pre-experience, work with free viewpoint videos, the notion of quality in general, the definition of QoE, and quality influencing factors and their detection were addressed as well. The analysis is qualitative and provides a summary, built on the experts' answers, following an Interpretive Phenomenological Analyis (IPA); see Coolican [99], p.233.

Expert Interviews: Results

The experts pointed out the importance of including subjective quality rating by the user in the development of an overall quality measure of 3DVO services. The experts provided a collection of factors, besides the Human Visual System (HVS) that could possibly influence quality: These factors are mainly based on a technical point of view. However, the experts agreed that these are probably not the only influences and that this should be verified by the particular user. They emphasized the lack of a *How to...* related to the integration of the particular users' opinion. Additionally, the experts'

color depth
compression
depth information
edge quality
realistic representation
resolution
synchronous texture
temporal stability

answers provided support for narrowing down possible pre-provided answers of the online questionnaire (see Section 4.2.1: Online Questionnaire).

Focus Groups

Two focus groups, after Morgan [159], were arranged. They received a brief introduction to the use case video communication, including the description of 3DVO technology (as described in Chapter 2). Therefore, a short explanatory presentation was necessary. The guideline used for conducting the sessions is presented in the Appendix A. Early adopters familiar with video communication took part. Participants had to fulfill four tasks and conclude with a discussion, in order to determine their usage of video communication, and wishes or ideas for further developments, including the presented new development. The tasks and questions are summarized in the following:

- Task 1: Description of video communication usage in the participant's day-to-day-life
- Task 2: Pair-wise simulation of a video communication situation noticing all factors that are important, useful, necessary, and desirable
- Task 3: Invention of a new video communication system which meets the requirements detected in task 2
- Task 4: Complementation of the described video communication usage in the day-to-day-life (from Task 1) with the new invention describing possible changes, added values, and other factors suggested by the participants.

These tasks were intended, in particular, to gather information on needs and requirements in video communication with 3DVO usage. The translated guideline for the conduct of the focus group and the questions asked can be found in Appendix A; originally, they were handed out in German. The analysis is based on the qualitative research approach, according to [159], and results in a description of the task outcomes and an outline of the consensus that the focus group defined in answering the research question(s).

Focus Groups: Results

The reactions of the test participants to the presented possibilities of free viewpoint video for video communication revealed an interest in eye contact and activity support. The consensus from the focus groups can be summarized as: *Wish for new developments supporting sociability and allowing device detachedness to a certain point, but without reducing the existing video communication system's quality, and providing a certain added value to the user.* However, the participants claim that their acceptance depends on already known quality factors (see results from the online questionnaire). The question on possible usage changes of video communication because of new possibilities was generally answered with a soft-spoken *yes*, because the participants think that their day-to-day life would not change. Similarly, their communication behavior would not change. However, wishes and visions for further developments concerning device detachedness and sociability support are pointed out e.g., in order to combine video communication with housework performed in parallel and information about user state.

Online Questionnaire

An online questionnaire (using Unipark [194]) was distributed to several participants without expertise in this particular field of research. General questions were asked on pre-experiences with general 3D technology and 3DVO in particular. Other questions aimed at detecting possible quality factors influencing a 3DVO representation. The expert interviews were taken into account when formulating them (questions are presented in Appendix A). The questions were worded so that they could not be answered with yes or no. This was grounded on the intention that already known parameters should be included and that additional important influences on the notion of quality are allowed. Subsequently, a selection of the provided questions is presented in the following. A translation of the complete online questionnaire, distributed in German, is given in Appendix A.

- With which applications using 3D technology do you have pre-experience? (Predefined answers were provided.)

- What pre-experience with video communication do you have?
- Do you prefer pre-defined displaying or rather the possibility of interaction and self-defined views? Why?
- How do you define the term *video quality*?
- In your opinion, which factors most strongly influence quality in relation to free viewpoint video quality?
- Which advantages and which disadvantages would free viewpoint video offer to you?

These questions do not focus on 3DVO in particular and not on the use case eye contact establishment by 3DVO. The main aim was to capture pre-experience and UX with free viewpoint video and video communication applications. Therefore, only possible applications were introduced explicitly, but not the TR process and 3DVO. The analysis is primarily descriptive and qualitative.

Online Questionnaire: Results

The average age of the 25 participants (17 men, 8 women) was 31 years. The findings show that 3D technology, especially 3DVO technology, is not very well known by the average participating user. When asked about pre-experience with 3D and video communication allowing multiple choices, participants only reported pre-experience with 3D cinema (21 persons out of 25). Other 3D applications, such as 3DTV (mentioned by 8 persons), User Generated Content (UGC) in 3D (1 person), virtual environments (e.g., CAVE) (8 persons), 3D PC games (5 persons), or 3D software (e.g., Maya, Computer-Aided Design (CAD)) (7 persons) did not receive much attention; neither did video communication (e.g., Skype, MSN) (6 persons). The majority (53 %) prefers pre-defined views, but, nevertheless, 47 % would like to choose the view actively on their own. As possible influences on quality were the already well-known factors related to the Two-Dimensional (2D) visual quality mentioned, such as: However, it is unclear whether they

color
contrast
depth impression
resolution

are based on pre-experience with common video applications or whether the participants really see them as the most important influence on free viewpoint video quality.

Summary of Explorative Approach

A summary of collected results is presented, based on the respective method(s) of analysis and a qualitative description. This summary is combining all methods used for the explorative approach and makes use of the richness of the collected data. This includes an analysis according to grounded theory, introduced by Glaser and Strauss [161], and text analysis, based on ten Kleij and Musters [195].

During the exploratory analysis, a list of presumed quality influencing factors and of influences on the acceptance of the perceived and produced quality of 3DVO was assembled qualitatively, taking into account the grounded theory approach according to [161]. The information is gathered from Literature Analysis A, the expert interviews, the focus groups, and the online questionnaires. Literature Analysis B was part of the exploratory analysis. As its results are more useful in terms of the proposed concept for an interlinking model, they are presented here and they are considered in Chapter 5. In this explorative approach, being comprehensive is not a requirement, but what is required is to reveal insights into which quality influencing factors may be considered as important.

The results are summarized as the following list of presumed quality influencing factors and influences on quality acceptance:

resolution
synchronicity
texture
edge quality (aliasing)
realistic representation
compression
depth information
depth impression
usability
temporal stability
contrast
color
support of social action

This list of factors could be divided into factors that are more related to the TR (bold) and factors related to perceptual domains only (italic). Presumed influencing factors that may lead to the links between single QEs and QFs were: depth information and depth impression, temporal stability and realistic representation, synchronicity, texture and realistic representation as well as depth impression. These findings have to be refined with further evaluation activities (see Study 2 and Study 3) and considered for the quality taxonomy model presented in Chapter 5.

4.2.2 Study 2: Laboratory Experiment

The laboratory evaluation was conducted in order to get information about the perceived quality of the new 3DVO and, especially, to find out more about possible quality influencing factors. Additionally, the influence of different usage context backgrounds (private conversation vs. professional conversation) was tested, in cooperation with the Institute for Psychology, University Salzburg, Austria.

Research Question

The research question of the laboratory evaluation was: *How do naïve participants perceive and describe different algorithms within the TR or 3DVO for the use case video communication?* This research interest includes the perception of eye contact, the perception of the overall quality, and the acceptance of the presented test material. Additionally, possible differences concerning these perceptions were tested exploratively, based on a fictional usage context (either private or professional conversation).

Method

The laboratory evaluation consisted of three parts: (a) A binary rating with yes/no, once related to quality acceptance and once concerning eye contact perception, as well as an ACR concerning the perceived quality, using a continuous rating scale between good to bad. (b) A qualitative, descriptive method asking for the description of the perceived quality in words. (c) A questionnaire about former media experience and experiences with video communication, including question related to social behavior, from the *Freiburger Persönlichkeitsinventar (Engl.: Freiburg Personality Inventory (FPI)*[196]. The test procedure is concluded with a brief discussion round to discuss open issues and answer questions from the test participants.

Table 4.3 summarizes the design of the evaluation study and shows the different predefined independent variables. These may be combined differently. The dependent variables were: *acceptability*, *quality judgment*, *quality description*, and *eye contact perception*. For the description of the different methods of synthesis analysis (S1, S2) and disparity analysis (A1, A2, A3), see Tables 2.3 and 2.4 in Section 2.3.2 and following subsection.

Table 4.3: Combination of the evaluation variables.

Technical characters of test items:	Usage context:	Content shown:
with/without eye contact	private conversation with friend	man A
two different views		woman A
different synthesis analysis	professional conference with adviser or consultant	man B
different disparity analysis		woman B

The test design used independent samples. This refers to the fact that the design was between-subject (i.e., different stimuli were judged by different people). Therefore, not every participant judged every test item depicted in Figure 4.2. This has the advantage of reducing the time needed per test participant to rate the provided test stimuli and of reducing order effects and memory effects. The disadvantage is the lack of variance homogeneity which leads to the usage of non-parametric tests. Hence, roughly equal numbers in each group of judges act as corrective (see Coolican [99], p. 75, and Table 3.3 for advantages and disadvantages of the various experimental designs).

Different conversational contexts (private conversation vs. professional conference) were examined. The consideration of these conversation contexts was realized with the usage of two differentiated mental models, according to [197], by asking the participants to project their thoughts onto one of the situations. This approach was used, because transmission and, therefore, real-time conversation, is not provided.

Test Material

The test items' contents were constituted by different 3DVOs showing four different people (two men, two women) resembling a possible video communication partner (see

Figure 4.2). Additionally, recordings from a pre-defined Ground Truth (GT) camera are shown for each content variant. These GT recordings show the ideal position and perspective the TR should reconstruct. The columns show the different TRs used and the GT recordings on the right hand side. The rows display the different contents employed with the stimuli, without horizontal correction (hc) on GT data, with eye contact at each third row. For the TR A2-S1, no men A without eye contact are shown (see first row in Figure 4.2). In total, 79 different videos were presented. A single clip had a duration of 10 seconds.

The test material was created using the creation process as described in Section 2.3.2. The five lines presented in Table 2.3 and Table 2.4 contain information written in bold face, describing the methods used for test item creation. For these evaluation activities, a set of test items was created using the methods described for disparity analysis and view synthesis. The settings differ in the combination of the various evaluation variables i.e., not every test participant had to rate all 79 videos (see test procedure and analysis). Instead, every test participant rated a complete set of stimuli with one of the four presented persons. For example, one test participant rated all available stimuli from man A and another participant rated the stimuli showing woman B.

A possible influence on the perceived quality by the displayed person was tested in a pre-test with GT stimuli only. Based on this, a different rating of the test stimuli because a person is attractive, inspires confidence, is likable, reliable, or perceived as being honest could be excluded. This exclusion is based on the pre-test results (see Appendix A). That there is no influence by attractiveness, etc. is true when considering the subjective perception perspective of this particular experiment. Nevertheless, non-excludable content relationships (i.e., based on the shown contents, as described in the section about test material), influence the rating, as will be demonstrated in the next section showing the results.

Test Environment

For every evaluation round, a group of a maximum of 5 participants met in the laboratory's anteroom and received the questionnaire, an introduction to the test procedure, and the task, as well as an introduction to the presentation and the Graphical User Interface (GUI)). Afterwards, each participant had a small room for the test procedure. The evaluation took place there, in a standardized environment i.e., lighting conditions

Figure 4.2: The different test items used.

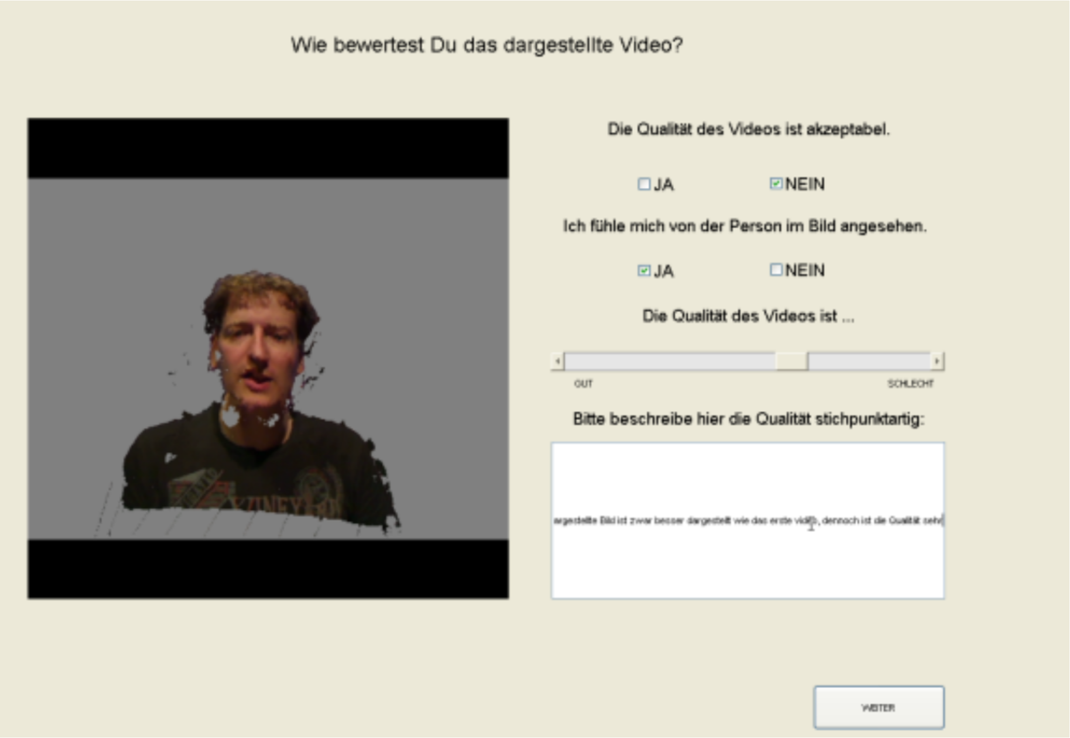
of 20 lux background illumination, with calibrated 19” LCD displays (see Figure 4.3). For the calibration a colorimeter (see, e.g., Spyder [198]) was used.

Figure 4.3: Small room with assessment GUI for individual participants.

Test Procedure

First of all, the participants filled out a questionnaire (paper and pencil) about demographic data, media experience, and social behavior, including questions from the FPI [196]. Subsequently, the small group of participants received a short introduction to the test including the explanation of the relevant mental model tested (i.e., private vs. professional conversation). Following this, each test participant entered one of the small rooms. Next, data sets on the perceived quality of used test videos were collected by a GUI created with MATLAB [199]. Herein, whether the test participant was asked in a rather private or in a professional way should support the respective tested mental model: Differences are represented by addressing the test participant either on a first name basis for a private conversation, e.g., see photograph of the GUI used for a private conversation situation Figure 4.4, or in a more formal way for a professional conversation.

Under these conditions, the participants were asked to decide whether the displayed video is acceptable or not (yes/no), to decide whether eye contact is perceived or not (yes/no), to rate the perceived quality on a scale (between left: good and perceptual directions (red lines): bad), and to describe the perceived quality in their own words (free text entry).



Wie bewertest Du das dargestellte Video?

Die Qualität des Videos ist akzeptabel.

☐ JA ☑ NEIN

Ich fühle mich von der Person im Bild angesehen.

☑ JA ☐ NEIN

Die Qualität des Videos ist ...

Bitte beschreibe hier die Qualität stichpunktartig:

argestellte Bild ist zwar besser dargestellt wie das erste video, dennoch ist die Qualität sehr

Figure 4.4: GUI private conversation context.

Finally, each participant returned to the anteroom for a final interview session, in order to discuss possible problems with the GUI and to obtain further information.

Test Participants

Participants were psychology students, between 17 and 38 years old (average age: 22 years); 260 were female, and 90 male. Because of the possibility to collect a total of 350 data sets, every possible test setting contains at least 15 data sets.

Method of Analysis

The data analysis was divided into three parts: Part (a) asked for overall quality, part (b) for the quality description in particular, and part (c) collected data about the test participants characteristics related to media and video communication usage and experience. For the analysis XlStat, a statistics package for Microsoft Excel, and R [200] were employed.

For the binary rating on quality acceptance, the perceived eye contact, and the ACR on quality are analyzed based on ITU-T BT.500 [36]. Additionally, the acceptance threshold after Jumisko-Pyykkö et al. [201] is calculated.

An exemplary definition of an eye contact-based threshold allows to draw more conclusions about the influence of eye contact perception on the perceived quality. It is calculated like the acceptance threshold introduced above, but using the binary eye contact perception rating instead of the acceptance rating.

For the qualitative, descriptive method asking for the description of the perceived quality in words, a cluster analysis (based on Glaser & Strauss [161]) is made.

For the data collected with a final questionnaire in the last part are analyzed with the procedure of descriptive analysis (frequencies, means, min./max. values, ranges, standard deviations, etc.), yielding a picture about the test participants' pre-experience with media in general and video communication in particular.

Results

The results of the laboratory evaluation reveal information about the perceived quality of the 3DVO created in different ways and about factors influencing quality. Overall, the results show that there are differences concerning eye contact perception and quality of the test stimuli. However, the main differences are made in terms of the shown content rather than between the different TRs. In the following, the results of the 3DVO evaluation are reported in three parts: Firstly, results of the psycho-perceptual rating of perceived quality, namely the acceptance of quality (yes/no), eye contact perception (yes/no), and the ACR, on a continuous rating scale between good and bad, as well as the collected quality description, are presented. Subsequently, results of the questionnaire about media usage and the questions based on the FPI are presented as well as conclusions from the open discussion at the end of the test procedure. Finally, this is concluded by describing the overall results with a more detailed investigation of the TR variants, the examined correlations, and a resumé that includes these separate parts. The analysis method was used as suggested in [99] and described in this section.

Results of the Psycho-perceptual Rating

All in all, the 3DVO presented did not provide an acceptable quality level for the test participants. On average, the GT presentation mode (without processing and scene representation) reached an acceptance level of almost 60 % and all produced 3DVO reached at least an acceptance of 10 %. With regard to the various creation processes, none of the test items (except GT) exceeded 13 % acceptance (see Figure 4.5). A Cochran's Q analysis to investigate differences between different presentation modes concerning the acceptance rating showed a highly significant difference to the GT presentation mode (Cochran's Q, critical: 12.59; observed: 839.06; $p < 0.0001$; $\alpha = 0.05$). This was supported by a post-hoc χ^2 test to test which of the presentation modes were actually different. Here, the acceptance rating showed a highly significant difference to the GT presentation mode: χ^2 (74, N = 3316) = 892.71, $p < .001$ too. The effect size was large, with Phi = 0.519. Based on this, it is also recognized that parameter combinations (i.e., different TRs, e.g., A1-S1 vs. A3-S2) are rated differently when averaged across content (displayed man and woman), taking into account the different stereo analysis and synthesis modes (A1-S1, A2-S1, A3-S1, A1-S2, A2-S2, A3-S2) with and without eye contact (ec) and hc, which was also presented in [19]. This means that it makes a difference whether the stereo analysis A1 or A2 is combined with the synthesis mode S1 or S2, for example. It seems that A2 and A3 combinations receive better ratings than A1 combinations, which finds expression only in combination with and across the different contents shown (see next paragraph on content, as well as results of the ACR between good and bad quality).

Focusing on the different contents (see rows in Figure 4.2) yields a similar result about the acceptance of quality. However, differences could be detected between contents. A Cochran's Q test of the differences between the contents and the acceptance rating showed a highly significant difference (Cochran's Q, critical: 14.07; observed: 200.79; $p < 0.0001$; $\alpha = 0.05$). A McNemar test confirmed this difference for the stimuli showing man A, which received a significantly worse rating, and the stimuli showing man B, which received a better rating. The reason for the bad rating for the stimuli with man A might be the extreme effect of the TR A1-S1, which led to the representation of a very long nose. This is supported by the quality description from the test participants mentioning this nose (see results of the collected quality description), an effect only found with stimuli showing man A. The reason for the better rating for the stimuli with man B might be the fact that even the stimuli with man B without eye contact showed a face more similar to eye contact, as the actor is sitting rather remote from the web

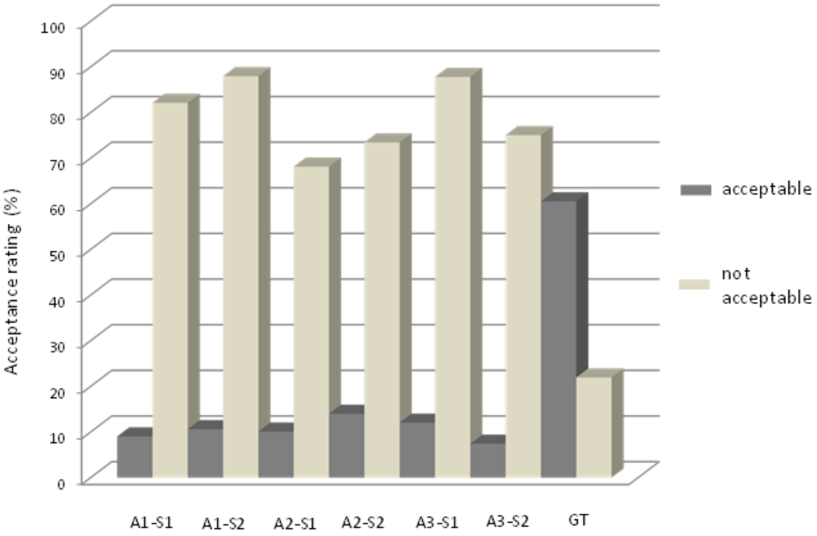


Figure 4.5: Acceptance of different Technical Realizations (TR).

cameras. This result is supported by the eye contact ratings, as presented below.

The eye contact ratings for all creation processes show similar results for the acceptance ratings.

Excluding the test stimuli without eye contact (see Figure 4.2, first row per man A, B and woman A, B each) leads to significant differences (χ^2 (19, N=30.14)= 1182.32, p<.0001. The effect size was large with Phi = 0.562.), as presented in Figure 4.6.

Paying attention to the different contents, it can be assumed that there are slight differences because of the content, but differences based on hc and eye contact differences outweigh possible content differences, which are probably also due to the ability of the TR to cope with the content e.g., differences because of the different hairstyles. Having excluded the GT presentation mode, no correlation between acceptance and perceived eye contact can be reported (see Figure 4.7). However, 19 % of the participants who perceived eye contact rated the overall quality as acceptable, and 9 % who did not perceive eye contact (average 13 %). Phi is positive but weak (0.152).

Results of the ACR between good and bad show that, here, the GT also differs from the ratings of the other stimuli created with different processes (Friedman's test, critical:

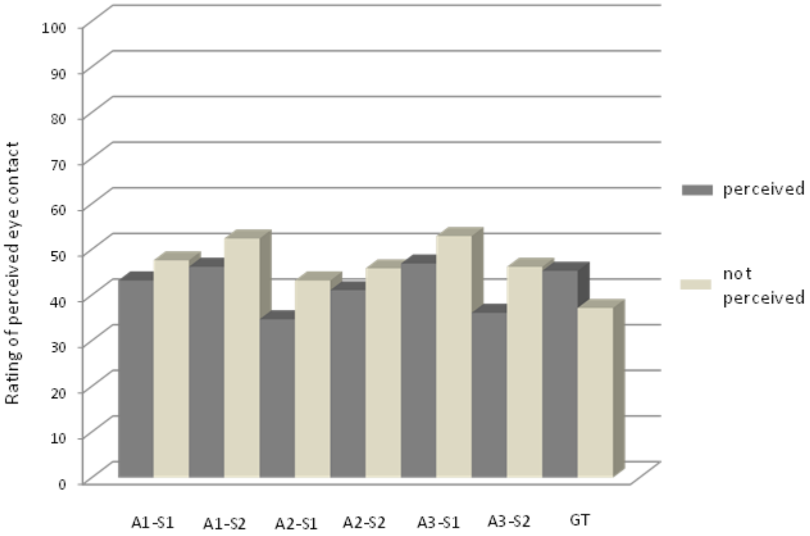


Figure 4.6: Eye contact perception of different technical realizations (TRs) without content showing no eye contact.

12.59; observed: 350.91; $p < 0.0001$; $\alpha = 0.05$), as shown in Figure 4.8 and Figure 4.9.

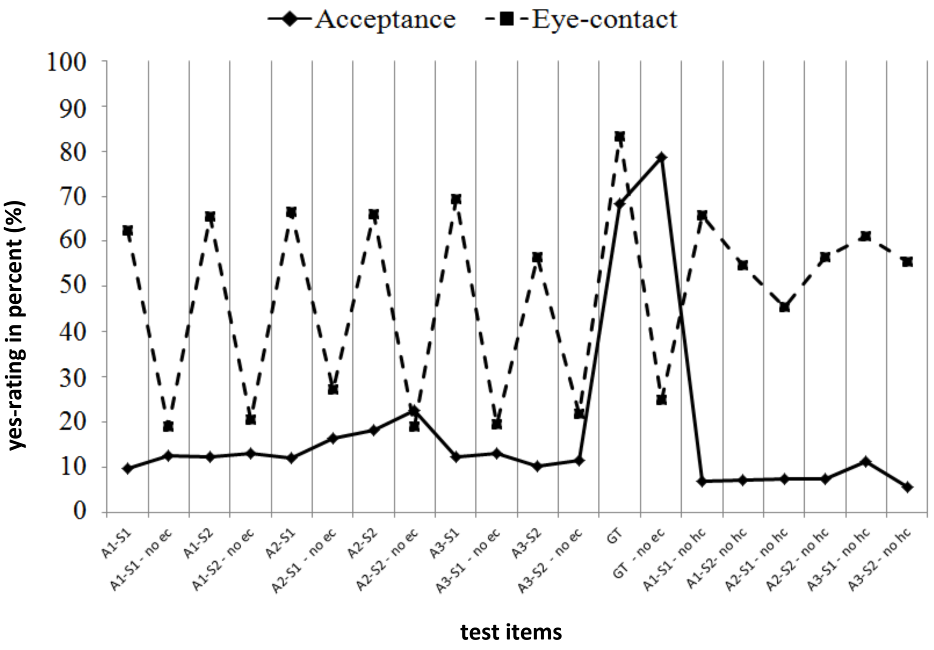


Figure 4.7: Acceptance of presented quality vs. eye contact perception in percent

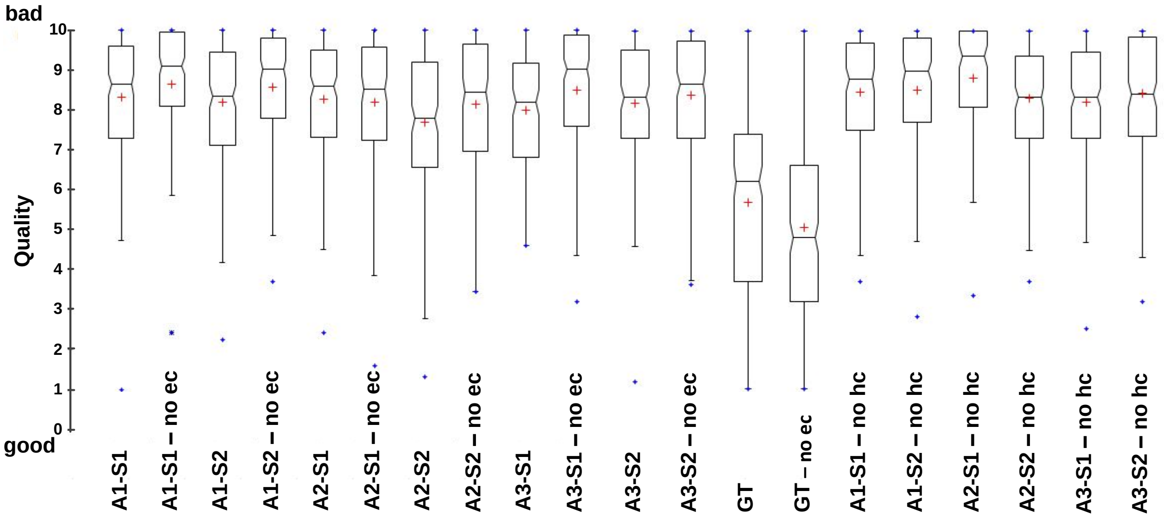


Figure 4.8: Quality perception of different algorithms with/without eye contact (ec) and horizontal correction (hc).

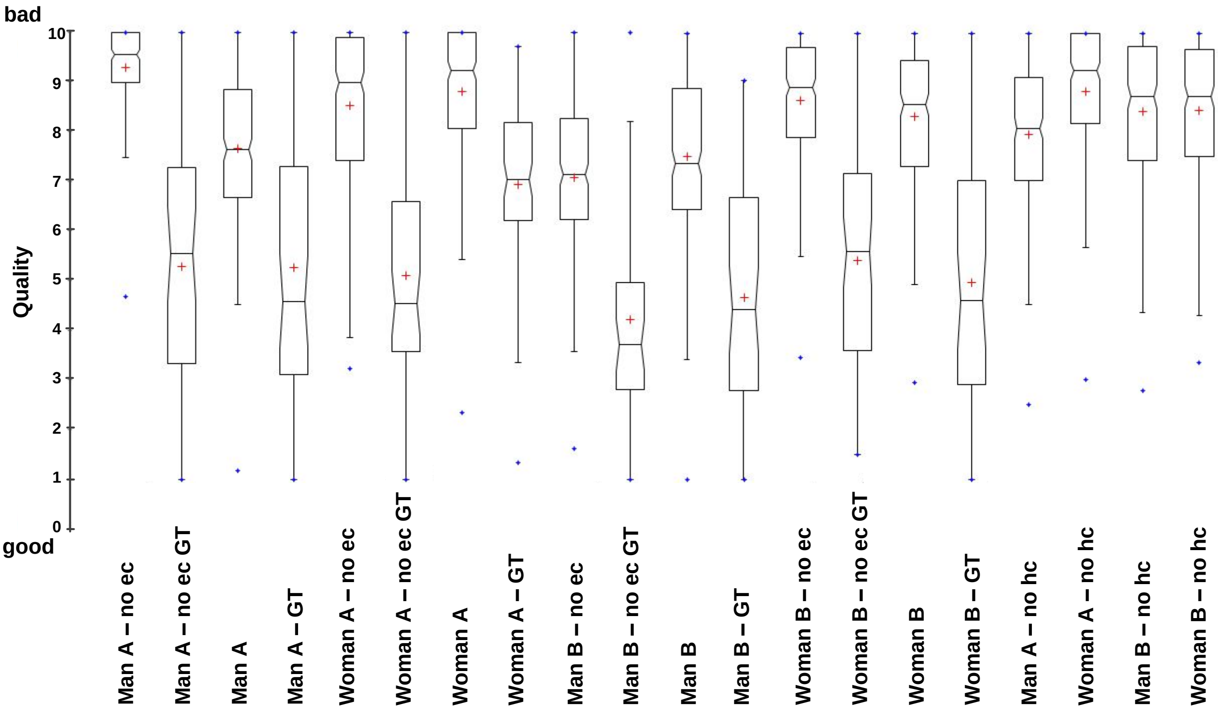


Figure 4.9: Quality perception of different contents with/without eye contact (ec) and horizontal correction (hc).

Unfortunately, compared to the other data sets, too few rating data are available for the man A stimulus created by TR A2-S1. Therefore, this data set was excluded from further analysis. Differences in quality ratings influenced by the different testing modalities and contextual tasks (i.e., talking to a friend vs. professional) could not be confirmed.

Ranking the available stimuli ratings based on the ACR results (quality rating between 1: good and 10: bad); excluding the stimuli with hc leads to the following order starting with the best assessment:

GT representations (SD: 2.30; min.: 1; max.: 10)

A2-S2 (SD: 1.72; min.: 1.33; max.: 10)

A3-S1 (SD: 1.49; min.: 3.22; max.: 10)

A3-S2 (SD: 1.56; min.: 1.17; max.: 10)

A1-S2 (SD: 1.43; min.: 2.25; max.: 10)

A2-S1 (SD: 1.60; min.: 1.61; max.: 10)

A1-S1 (SD: 1.57; min.: 1; max.: 10)

This result is especially interesting, when compared to the result of man A (i.e., long nose) created with the above-mentioned TR A1-S1. This leads to the assumption that quality ratings in the low quality range can be easily differentiated, even though the overall quality may be rated as very low. Furthermore, it can be supposed that this particular content leads to different results when focusing on content as the influencing factor, rather than on the overall quality or on a special influencing factor, as was the case in the pre-test presented in Appendix A. However, these assumptions should be investigated in future activities.

Results of the Collected Quality Description

Analyzing the individual qualitative descriptions of the perceived quality of each participant, taking into account the grounded theory approach, according to [161], and the frequency of the same descriptive meanings, leads to the following list of quality influencing factors (presented from high (1.) to low (11.) frequency). It includes the most frequently mentioned quality influencing factors and possible influences on quality acceptance.

1. incomplete
2. contortions
3. lagged
4. blurry, not sharp
5. pixelated
6. blemished
7. shape distortions
8. jitter
9. wrong proportions
10. unlikely background
11. bad color quality

The constraint of this form of analysis is the influence of the interpreter who decides whether a mentioned quality description has the same meaning as another or not. As seen here, some categories may be part of another. In order to overcome this particular obstacle, the OPQ study was applied in a further step. Another difficulty is to simply translate attributes from German into English. For this case, a clear definition of each single defined attribute was noted and then translated into English, as listed in Appendix A in Table 6.3 on page 204.

Results of the Questionnaire and FPI

Overall, 132 data sets useful for further analysis were available. The reasons why so many data sets from the more than 300 participants had to be excluded are twofold: First, incorrectly filled out questionnaires reduced the data. Second, conflicts with the test participant number and the ratings provided by the GUI that could not be synchronized afterwards again by verifying with additional desktop recordings of each conducted test were excluded. The analyzed data consist of data sets from 33 men and 99 women with an average age of 22 (ages between 18 and 35 years). Almost half of the participants (60) were using glasses or contact lenses. Only 3 of the considered participants reported that they are color blind. On a Likert scale (i.e., the scale contains equal numbers of positive and negative positions and provides same distances between each position) of 7 points (1: very familiar, 7: not familiar at all), the majority (116) indicated that they felt quite familiar with new media (51 ticked off point 3, 26 point 2, 23 point 4, 16 point 5), followed by feeling rather unfamiliar and feeling very familiar (7 ticked off point 1). With regard to the most frequently used kind of media (between 1: several times per day, 2: daily, 3: several times per week, 4: weekly, 5: monthly, 6: more seldom; 7: never) the internet (109 ticked off 1) followed by the cell phone (93 ticked of 1) are the most frequently used kinds of media. Video communication is not used very often; the majority (28) reported that they never use video communication. However, 26 seldom use video communication, 23 use it weekly, 22 use it several times per week, 20 use it monthly, and only one uses it several times a day. The answers to the FPI show that the majority of the participants is rather sociable and extroverted. There was no dataset with a contradictory result.

Summary of Results of the Laboratory Analysis

The results of Study 2 reveal that, although the quality rating took place in the low quality range and content seems to play an influencing role, distinctions between the different TRs used are made (when combining different TRs and content). This is not only shown in the objective data analysis (see part (a)). When focusing on the correlation between eye contact perception and quality acceptance, other factors influencing quality also need to be considered in future data analysis activities. This may include a comparison of the data analysis in part (a) with quality describing vocabulary per investigated TR. Comparing the quality describing vocabulary from part (b) with the collected information on quality influencing factors or, especially on quality influencing QFs in Study 1, shows that the new representation and the different kinds of TRs produce quality influencing factors other than those considered until now. This is especially true for factors such as incompleteness (holes etc.), contortions, lagged parts, and blemished areas, because these are TR-specific factors explicitly including temporal effects in addition to preliminary findings. However, the cluster analysis of the vocabulary is highly influenced by interpretation activities of the investigator. Therefore, a further study (Study 3) should help to let discriminate QFs already correlated with each investigated TR by the user. Subsequently, a comparison of the description of used TRs and considered problems (see Section 2.3.2) with the presented evaluation results is investigated and considered for the extended quality taxonomy definition in Section 5.3.

4.2.3 Study 3: Open Profiling of Quality (OPQ)

The goal of the second laboratory evaluation was to obtain more information about perceived quality factors related to 3DVO from different TRs. Quantifying these quality influencing factors was an additional aim. Therefore, the OPQ method was chosen. The advantage is that attributes describing quality are developed and rated by the test participants, thus excluding an interpretation task for the investigator (i.e., the clustering of quality description as applied in Section 4.2.2). In doing so, it was again intended to profile users' perception of different 3DVOs developed for the video communication application, as a step towards the definition of QFs.

Research Question

The overall question was whether different 3DVOs based on different TRs are differentiated in subjective quality assessments and descriptions. In detail, it was asked whether the test participants accept the presented 3DVO, whether they perceive eye contact, and rate the perceived quality differently, based on the various algorithms used for the TRs. Again, the question of *why* the quality was rated in that way was addressed by asking for a description of perceived quality and its quantification. This leads to the definition of QoE influencing QFs.

Method

OPQ intends to construct a deeper understanding on subjective quality assessment (see Strohmeier et al. [157]). Thus, this method allows the combination of a qualitative descriptive quality evaluation, based on an individual's own vocabulary, with a quantitative psycho-perceptual evaluation, based on standards (i.e., ITU-T BT.500 [36]). It is a mixed method approach that targets naïve participants in experiments with miscellaneous stimulus material for multimedia quality evaluation. For a detailed description of the sensory profiling approach of this method, see Strohmeier et al. [157] and, for the applicability of the sensory evaluation, see Section 4.2.1.

Test Material

The test included 44 video clips showing a speaker who represented a possible conversation partner. Each video clip lasted 10 seconds. These test items were presented in 6 TR combination variants and 4 content variants. The content varied by showing different persons: three men and one woman. Out of 44 test items, 4 processed 3DVOs were not horizontally corrected. These were included in the perceptual evaluation, but not in the sensory profiling task. The 6 TR combination variants contain the following stereo analysis/synthesis combinations: A2-S3, A4-S4, A4-S5, A5-S4, A5-S5 and an A2-S3 combination without a horizontal correction (hc), which means the person shown was not returned to the initial recorded position and therefore seems to hover (see last row in Figure 4.10 and description about hc in Section 2.3.2). The first 5 combinations were shown once, including an artificial background, and once without a background, showing only the person as recorded. Additionally, a variation concerning eye contact was recorded. One of the recorded men does not show any eye contact (see Figure 4.10).

In Figure 4.10, the four content variants are depicted in the four columns and the TR variants are represented in the rows. The third column shows the man without eye contact. The various contents are shown in different frame sizes, which is due to the hc and is seen as part of the respective TR. A possible influence on the subjective assessment is analyzed. The test items were created with a slightly modified TR, compared to the laboratory study (Study 2). This modification is based on the further development of the TR chain. Table 2.3 and Table 2.4 (in Section 2.3.2) present the different creation processes within the laboratory experiment, using OPQ. The six lines written in italics contain the methods used. The differences between the TR variants concern the disparity estimation, the post-processing, and the virtual view synthesis. Additional differentiations regarding aggregation and the pre-processing of the video material for the stereo analysis are made: The variants A2 and A4 were created using a Gaussian approach for the aggregation and were pre-processed with a noise reduction. Variant A5 has no aggregation and no pre-processing. All the test items were created by applying rectified grey scale monochrome input as the source for the stereo analysis. All test items were prepared for the synthesis by using rectified information and depth information from the left camera. These stereo analysis and synthesis processes are described in more detail in [19] and [1].

Figure 4.10: Columns showing different contents and rows showing different technical realizations TRs (top down: A2-S3, A4-S4, A4-S5, A5-S4, A5-S5, A2-S3-BG, A4-S4-BG, A4-S5-BG, A5-S4-BG, A5-S5-BG, and A2-S3 no hc).

Test Environment

The test items were presented with a duration of 10 seconds each, a resolution of 640x480, and without audio. The test items were shown in randomized order. The controlled laboratory conditions were similar to the laboratory environment (i.e., standardized lighting conditions based on recommendations), but without an explicit (constructed) conversation context. The context was a controlled laboratory environment at the Technischen Universität Ilmenau, Germany. The test conditions were arranged according to the standardization requirements of ITU-T BT.500 [36] and ITU-T P.910 [141]. This included the viewing distance for the participants, the ambient light conditions, the lighting of the display and the background color presented on the LCD and within the GUI.

The presentation took place on a 19” LCD display embedded in a GUI and based on the original intention to resemble a video conversation tool (see Figure 4.11), where the conversation window of a certain size is placed in a certain area on the display. This GUI was created in order to replace paper-and-pencil answers to the questions.

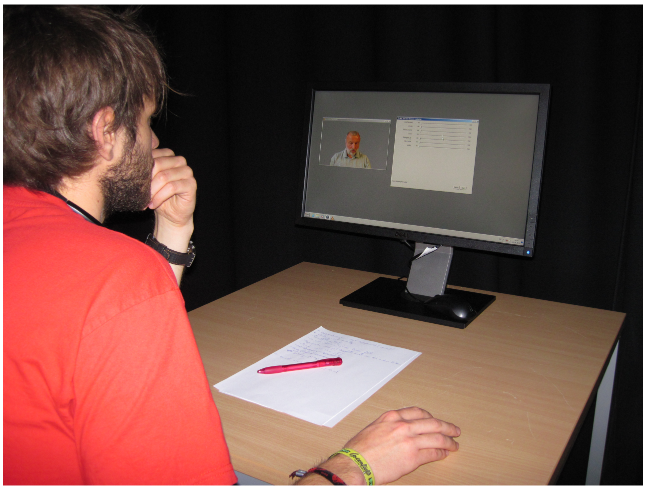


Figure 4.11: Test participant in front of a 19” display in the laboratory environment.

Test Procedure

The quantitative psycho-perceptual evaluation and qualitative descriptive task (sensory profiling) was performed with all participants. The participants were asked for their assessment of randomly presented 3DVOs. The assessment concerned the acceptance of displayed quality (yes/no), the rating of the displayed quality (between good and bad), and its rating and the perception of eye contact (yes/no) in two separate, quantitative parts. In between these two parts, the sensory evaluation took place, in which participants generated their individual quality describing attributes, based on [157], and rated them. The test was performed once within 90 minutes and divided into three main parts. It started with a visual screening of the participants and continued with (a) a quantitative part that asked about general quality acceptance and perceived overall quality. The second part (b) was dedicated to the sensory profiling, asking for individual quality describing attributes and their occurrence, as suggested in [157]. The third part (c) consisted of another quantitative aspect: focusing on eye contact perception. Finally, a questionnaire (d) concluded the test procedure.

With the help of a visual screening, the participants initially received an explanation of the test procedure. As part of a preliminary training, a subset of the test items covering the full range of quality was presented, in order to allow participants to familiarize themselves with the presentation mode.

Psycho-perceptual Evaluation

At this stage of the test procedure, the test participants practiced the evaluation task. Then, they evaluated the quality of presented stimuli quantitatively. The stimuli were presented successively, in random order, and the participants had to rate their acceptance of the quality on a binary scale, and the overall satisfaction on an unlabeled 11-point scale, as suggested in [157].

Sensory Evaluation

The second part started with an introduction into the sensory evaluation task. The aim of this evaluation was to offer participants the opportunity to generate their individual descriptive vocabulary. Therefore, the exclusion of evaluators' influences and the exclusion of a possible pre-definition of describing attributes are mandatory. Hence, each participant had to develop an individual vocabulary without any suggestions from the evaluator. For this reason, the first task contained the so-called *apple task*, as suggested in [157], to give the participants the opportunity to become familiar with the sensory evaluation procedure and the task of defining a describing vocabulary. This supports the participant in obtaining a better understanding of the evaluation goal and to become familiar with the procedure itself. Subsequently, the participants developed their individual quality attributes while viewing a representative subset of the test stimuli. After a refinement of the quality attributes, including a clear definition of each attribute and its extremes, the final individual attribute list for quality rating was specified. This specified attribute list was presented with an 11-point scale labeled with min. and max. per attribute, with min. denoting that the attribute is not perceived at all, and max. referring to its maximum extent. This was presented within the GUI for each presented stimulus. The participants rated the presence of each individually pre-defined attribute for the test items presented in random order.

Perceptual Evaluation of Eye Contact

The third part of the test procedure consisted of the task of rating whether eye contact with the shown person was perceived or not. Here, all 44 test items were presented, again in random order, and the participants had to decide if eye contact is perceived or not, on a binary scale (yes/no).

Questionnaire

In a final step, the participants filled out a questionnaire asking for demographic data, viewing conditions (categories), media (pre-)experience (7-point scales), and social behavior, including questions from the FPI [196].

Test Participants

In sum, 32 naïve participants took part in the study (18 male, 14 female, ages: 19–53; average age: 25). They were performing both the quantitative part (psycho-perceptual evaluation) and the sensory profiling. The majority (22) of the test participants indicated that they uses optical aids (e.g., contact lenses, glasses). None of the participants were color blind.

Method of Analysis

For the data analysis, XlStat and R were used, as in the previous laboratory experiment.

For the psycho-perceptual evaluation and the perceptual evaluation of eye contact, the data collected from a binary rating on acceptance (yes/no), eye contact perception (yes/no), and from an ACR (11-point scale) for the overall quality rating, were analyzed, based on ITU-T BT.500 [36] and used for the definition of an acceptance threshold, based on [201].

Therefore, after a normality check for the data and a descriptive analysis, significance tests were conducted, as suggested in [104]. These provide information about the differences in quality perceived by the test participants. Moreover, the acceptance threshold according to Jumisko-Pyykkö [201] was defined.

For the sensory evaluation, the data were used to conduct a GPA and a preference mapping, based on the OPQ procedure of Strohmeier [104]. The GPA is applied in order to find out how many components were needed to explain a 100 % variance in the resulting model, and which components are useful for further data interpretation. This allows an interpretation of the consensus of different test participants for the relevant quality presented. To do so, a further interpretation of the GPA results, after Lorho [151], can be carried out for the quality assessment of the contents as well, but mainly for the TR algorithm combinations used (see Figure 4.10). Additionally, a separate Principal Component Analysis (PCA) may lead to further conclusions, including the creation of confidence regions (e.g., marked by ellipses), as introduced by Husson and Pagès [202], in order to provide a graphical presentation of the significance of differences between the evaluated TRs.

The data collected with a final questionnaire in the last part was analyzed with the usual procedure of descriptive analysis (frequencies, means, min./max. values, ranges, standard deviations, etc.) to reveal a picture of the test participants' pre-experience with media in general and video communication in particular.

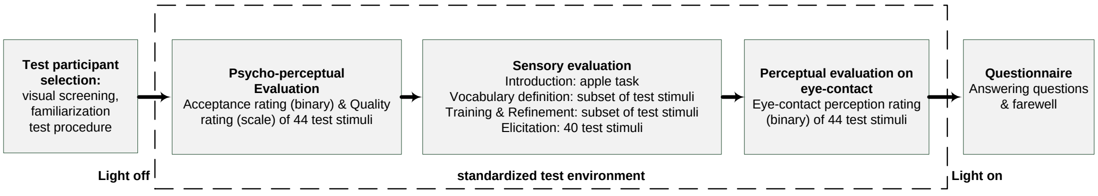


Figure 4.12: Overview of the OPQ procedure used (after Naes and Risvik [130]).

To summarize the applied OPQ method, Figure 4.12 provides an overview of the overall test procedure, the tasks, and the focus for each part, as described in this section.

OPQ Results

The results of the OPQ are based on a quantitative analysis and a sensory analysis, according to the description in [157]. In addition to conventional descriptive data, results about the participants' media experience and social behavior (i.e., level of intrinsic/extrinsic personality and openness), based on the FPI [196] questions, are presented. Generally speaking, the results highlight the importance of taking into account differences in contents when testing for differences caused by different TR processes. There is no TR process which was rated extraordinarily better or worse than the rest. However, small differences related to the TRs used could be detected after checking for differences in the different contents that were used both in the quantitative rating and in the sensoric quality description. The results of the quantitative analysis show that a differentiation of the overall quality is made, in line with the depicted content. This, although the possible differentiation of different contents (i.e., conversation partners) could be neglected in the test stimuli when investigating attributes like beauty, trustworthiness, honesty, etc. in a previously mentioned study. In a pre-test for Study 2, the laboratory experiment and all framework conditions (i.e., T-shirt worn, background, camera position) remained the same (see Appendix A). Results showed no significant difference whether the test stimuli contained background or not. Inspecting a look on the influence of eye contact perception reveals a trend regarding a connection between perceived eye contact (frequency, in percent, of all yes ratings) and quality acceptance (frequency, in percent, of all yes ratings) as shown in Figure 4.13.

In the sensory data analysis, the most substantial quality describing attributes collected were: missing picture parts (holes) and flicker or jitter, which influences the perceived quality. The attributes generated, in German, and relevant for interpretation (i.e., they have a coefficient of determination (R^2) between 0.5 and 1) are translated into English for further presentation in this work and can be found in Appendix A - evaluation material (see Table 6.3 on page 204).

In the following, the detailed outcomes resulting from the 3DVO evaluation are reported in three parts: Results (a) from the psycho-perceptual rating of perceived quality, (b) from the sensory evaluation, and (c) of eye contact perception. Finally, (d) results of the questionnaire about media usage and answers to the questions based on the FPI are presented. This is concluded by considering the overall results, with more emphasis

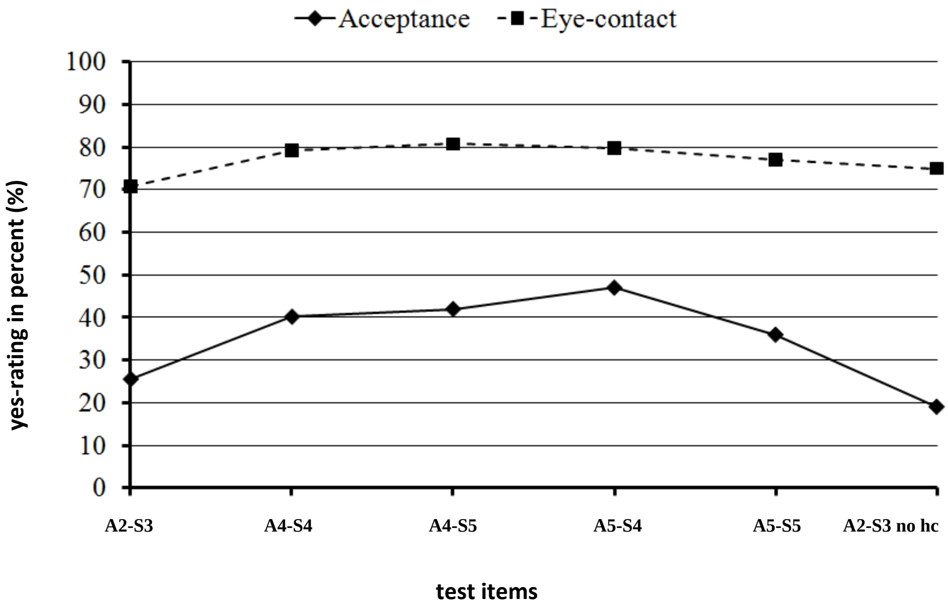


Figure 4.13: Acceptance of presented quality vs. eye contact perception, in percent.

on the TR variants and a resumé that includes these separate parts. The method of analysis was applied as suggested in [157] and described in Section 4.2.3.

Results of the Psycho-perceptual Rating

The presented test items did not provide a quality level acceptable to the participants. Overall, 40.5 % rated the presented quality as acceptable. With regard to all of the variants (i.e., content and different TR combinations, as shown in Figure 4.10), none of the test items was rated below 3.2 % and above 90.5 % of acceptance. The tendency is towards a relatively low level of quality, with 37.3 %, taking the median as a measure of the central tendency for ratings across all variants.

Inspection of the stereo analysis and synthesis modes (A2-S3, A4-S4, A4-S5, A5-S4, A5-S5, and A2-S3 without hc) reveals that TR combinations seem to have an influence on the overall quality satisfaction, when averaged over content (displayed men and woman, with and without artificial background). These influences on acceptance have been further investigated. The results are presented in Figure 4.14.

Figure 4.14 shows differences perceived by the test participants, based on the different TRs used (excluding A2-S3 without hc) for the 3DVO. Some variations are rated signif-

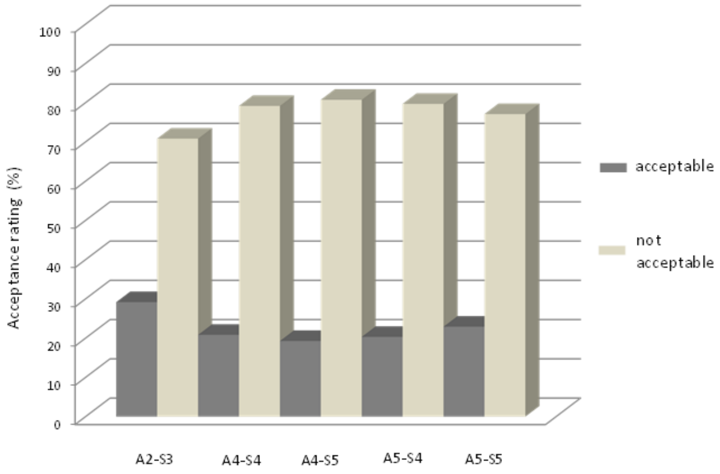


Figure 4.14: Acceptance rating for different TR combinations.

icantly different from other stereo analysis and synthesis mode combinations (Cochran's Q, critical: 9.49; observed: 174.62; $p < 0.0001$; $\alpha = 0.05$). In particular, TR combinations A2-S3 to A4-S4, A2-S3 to A5-S4, and A4-S4 to A4-S5, A4-S4 to A5-S5 (McNemar: $p < 0.0001$), as well as A2-S3 to A4-S5 (McNemar: $p < 0.002$), and A4-S4 to A5-S4 (McNemar: $p < 0.003$) are rated significantly different. Differences in acceptance ratings (i.e., acceptance in percent, based on yes/no frequencies) of the TR A2-S3 with and without hc could be detected (Cochran's Q, critical: 3.84; observed: 6.33; $p < 0.012$; $\alpha = 0.05$). However, no significant differences could be located when using the McNemar test for pairwise comparison (McNemar: $p = 0.36$).

With regard to the quality rating based on the scale between good and bad, Figure 4.15) shows the results. A normal distribution of the ratings of all TR variants is not present. Most of the variants have significant differences (Friedman, critical: 9.49; observed: 276.80; df: 4, $p < 0.0001$; $\alpha = 0.05$) except TR A4-S5 to A5-S4 ($p = 0.24$).

Differences in quality ratings (i.e., ACR between good and bad) of the TR A2-S3 with and without hc were detected (Friedman, critical: 3.84; observed: 8.81; df: 1, $p < 0.003$; $\alpha 0.05$) and these are assumed to be significant (pairwise difference, $p = 0.02$; $\alpha = 0.05$).

Combining the results of all TR variants concerning quality acceptance and quality rating

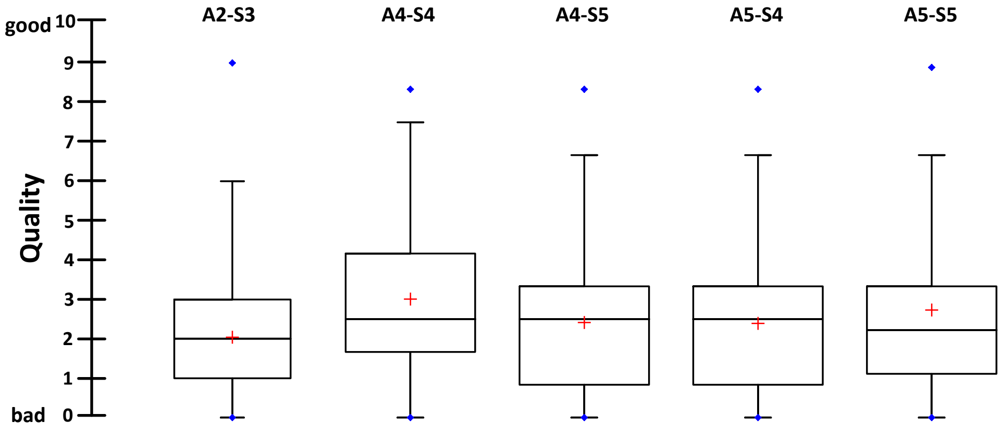


Figure 4.15: Quality rating of various TR combinations.

between good and bad, yielded the following definition of threshold: The scores for the unaccepted overall quality lie between 6 and 0 (mean: 1.86, SD: 1.33). Accepted quality was rated between 0 and 10 (mean: 4.04, SD: 1.78). Thus, no acceptance threshold could be determined, because indicators relevant for threshold were detected between 6 and **0**. This is also the case when analyzing acceptance and quality rating for each TR separately.

Results of the Sensory Evaluation

In the second part of the evaluation, the participants developed a total of 199 individual quality attributes for sensory analysis. Although two of the attributes were rated in a binary style (only investigating max. or min. with nothing in between) and the attribute "long-drawn-out" was always rated with min. and, hence, non-existent, they were retained for further processing. They need to be kept in mind for further interpretation activities. The GPA was applied, as suggested in [157] and described in [203], in order to determine how many components are needed to explain 100 % variance in the model, and which components are useful for further interpretation. In this case, 9 components are needed as shown in Table 4.4. In the following, only the first two components are considered for each presented TR result, because the other components together do not exhibit more than 50 % accountancy (see row 2, Variance, in Table 4.4).

Table 4.4: Eigenvalues, variance, and cumulative variance of components used to explain 100 % variance.

Components	F1	F2	F3	F4	F5	F6	F7	F8	F9
Eigenvalue	704.852	512.327	386.550	136.101	96.930	21.847	9.441	7.441	3.050
Variance (%)	37.521	27.273	20.577	7.245	5.160	1.163	0.503	0.396	0.162
Cumulative variance %	37.521	64.794	85.371	92.616	97.776	98.939	99.442	99.838	100.00

Figure 4.16 shows a correlation plot that presents the first two components (F1 and F2), which include factors mentioned by all participants (K1-K32) and account for 64.79 % of the explained variance (dimension 1, F1: 37.52 %, dimension 2, F2: 27.27 %). These two dimensions form the perceptual space. Attributes with coefficients of determination (R^2) between 100 % and 50 % are emphasized and considered for further interpretation.

Figure 4.16 shows these attributes, with representatives written in color (different colors denote the attributes per test participant), and close-to-borderline ratings written in grey. The correlation plot shows attributes written in English. However, because the attribute elicitation was carried out by German native speakers, a simple translation may not be appropriate. Hence, for the following textual description of the results, a carefully considered translation was used. A table with the translated attributes is given in Appendix A. The positive polarity of dimension 1 is described with attributes such as stripes, flicker representation, and fringed edges. Its negative polarity is described with attributes more related to missing content, such as holes and transparency. For Dimension 2, the positive polarity is described with attributes related to flicker and noise as well as to holes, transparency, grey spots, and aliasing, whereas the negative polarity is described with attributes linked to completeness and color of the shown picture. It can be noticed that the positive polarity of both dimensions is stronger. The attributes positioned at the highest polarity, with more than 75 %, are about holes, missing parts of the picture, disturbing pixel lines, lines, and stripes. After this general overview on the perceptual space, taking into account all valuations of all presented test items perceived by all test participants, a closer look at possible differences between the different creation variants is given in the following (see Figures 4.17 to 4.21), which includes developed attributes from all participants (K1-K32). As shown in Figures 4.17

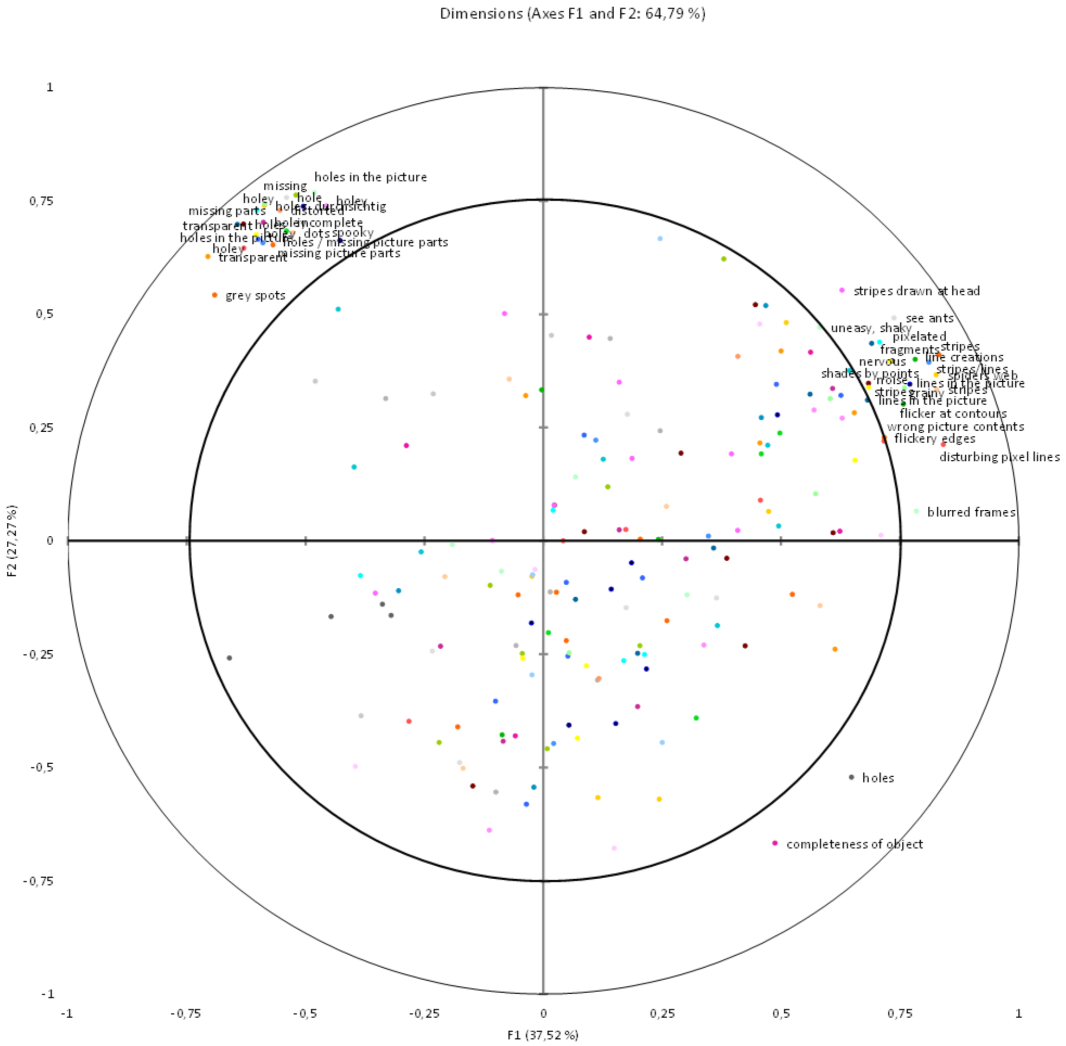


Figure 4.16: Correlation circle of attribute ratings for all stimuli showing the first two dimensions F1 and F2.

to 4.21, the perceived quality of the different TRs are described differently:

For A2-S3, the first two components (F1 and F2) account for 79.19 % of the explained variance (dimension 1, F1: 50.74 %, dimension 2, F2: 28.45 %). These two dimensions form the perceptual space. As shown in Figure 4.17, attributes with a R^2 between 100 % and 50 % are mainly related to the completeness of the picture. This is represented by attributes at the positive polarity of F2 and at both polarities of F1, which are, for example, hole, flicker, color/grey spots, wrong color, deformations, and blurring. The attributes related to the aspects background and completeness at the positive polarity of F1 support the assumption about holes as the greatest problem of TR A2-S3.

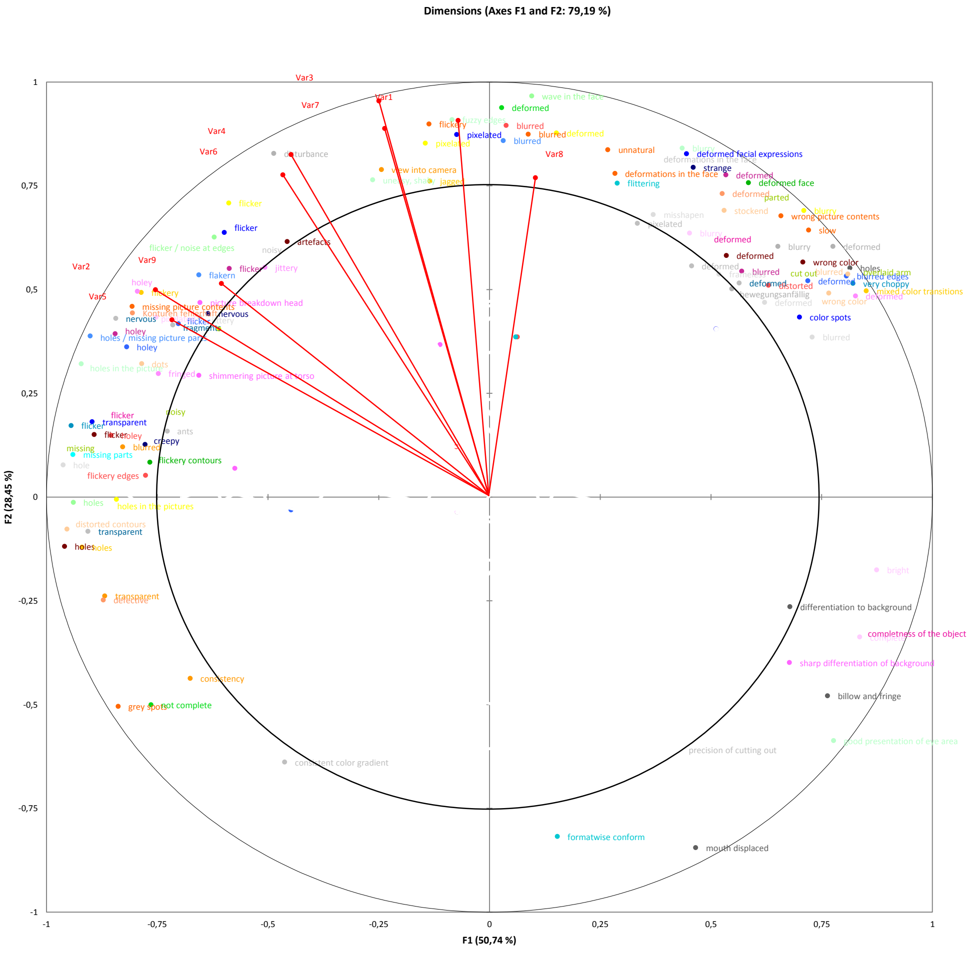


Figure 4.17: Perceptual space and perceptual directions (red lines) of TR A2-S3.

Taking into account the *perceptual direction*, one can see that the perceived severity of holes, flicker, and blur are represented by perceptual directions, whereas deformations and wrong color seem to be less severe. This consideration of perceptual directions is suggested in Naes and Risvik [130]. It is similar to the definition of Lorho [152] but excluding an additional PCA conducted with the group average data derived from the presented GPA. The results derive from a comparison of the perceptual space and the perceptual directions (red lines) of Figure 4.17.

For A4-S4, the first two components account for 70.91 % of the explained variance (dimension 1, F1: 46.68 %, dimension 2, F2: 24.24 %). The perceptual space presented in Figure 4.18 contains attributes with a R^2 between 100 % and 50 %, such as contorted, lines, fragments, pixelated, blocky, very choppy, slow, nervous, and picture breakdowns. In particular, the positive polarity of F1 is described with these attributes. Attributes such as lines, noise, ants, flicker, stripes, as well as blur, pixelated and fragments define the negative polarisation of F2. The attributes with positive polarization of F2 are related to the background and to positive messages, like steady color gradients, format-wise conform, eye parts well presented. TR A4-S4 seems to have more problems with outliers than with missing parts, compared to TR A2-S3.

Interpreting the perceptual directions by comparing the information presented in Figure 4.18 leads to the assumption that aspects such as noise, blur, fragments, pixelations, unnatural representations, and picture breakdowns in the head and face parts of the picture are the most severe problems with TR A4-S4.

The perceptual space for A4-S5 accounts for 75.40 % of the explained variance formed by the two dimensions F1 (46.92 %) and F2 (28.48 %), as presented in Figure 4.19. Among them, the attributes worth mentioning and with consideration of the determination coefficient are blurry, pixelated, contortions, distortions, noise, flickery edges, fuzzy edges, distortions in the face, stripes, blocky, overlaid arm, and color distortions. These attributes are part of the positive polarization of F2. The attribute holes is mentioned once at the positive polarization of F1. This polarization at the negative polarization of F2 also contains compression artifacts and positive attributes like format-wise conform, steady color gradients, equability, and eye parts well presented.

The perceptual directions seen as red lines in Figure 4.19 lead to blur, picture breakdowns in the torso area, stripes, contortions, and flicker, as the most severe quality attributes.

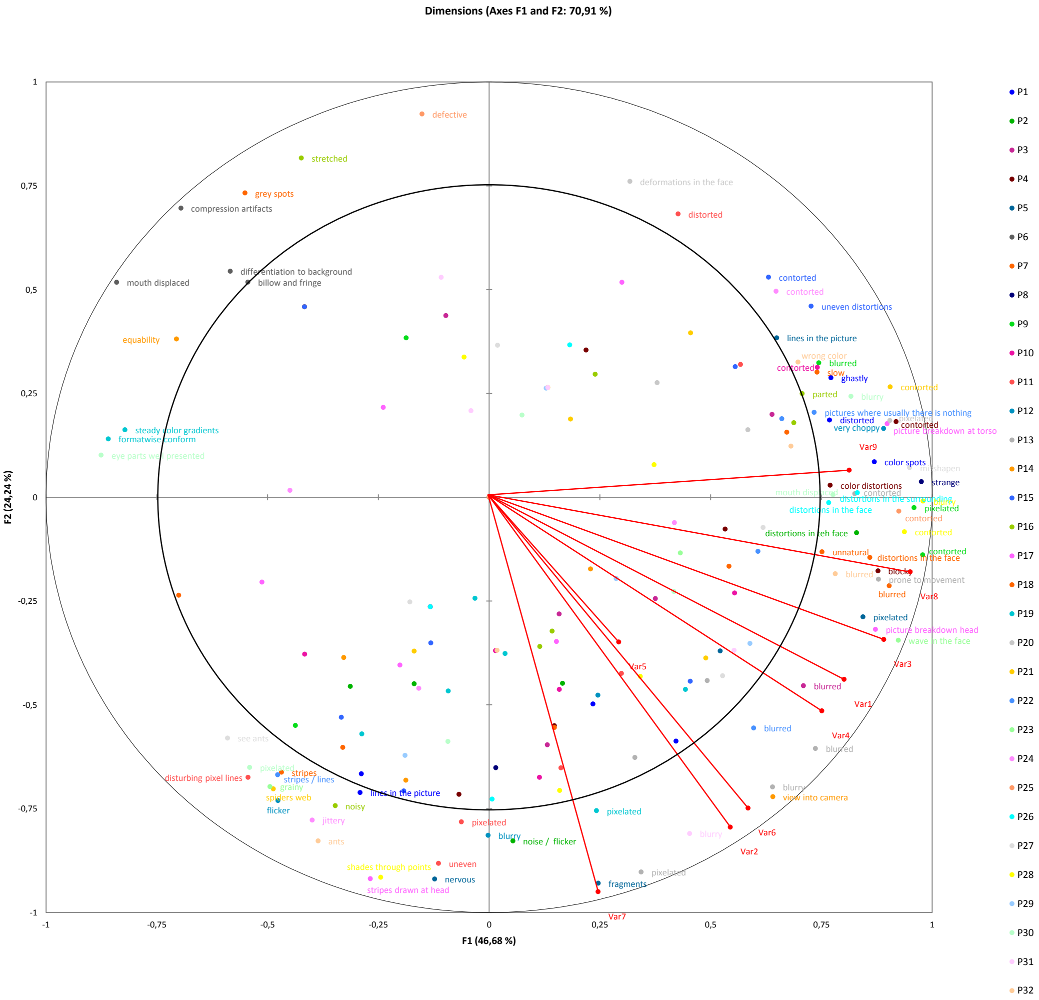


Figure 4.18: Perceptual space and perceptual directions (red lines) of TR A4-S4.

As presented in Figure 4.20, A5-S4 is described by a perceptual space accounting for 67.42 % of the explained variance (dimension 1, F1: 46.85 %, dimension 2, F2: 20.57 %) and especially provides stripes, noise, pixelated, nervous, flicker, blurry, deformations, jitter, and lines in the picture as attributes exceeding the coefficient of determination. Thus, the positive and negative polarity of F1 and F2 contains quality describing attributes rather related to blur, additional but wrong picture content (i.e., lines, pixels, stripes, grey spots, ants, noise), and deformations. The attribute holey is mentioned relatively seldom, but it seems to be a problem (e.g., attributes such as picture breakdown). However, the incorrect assignment of picture parts seems to be the major problem of TR A5-S4.

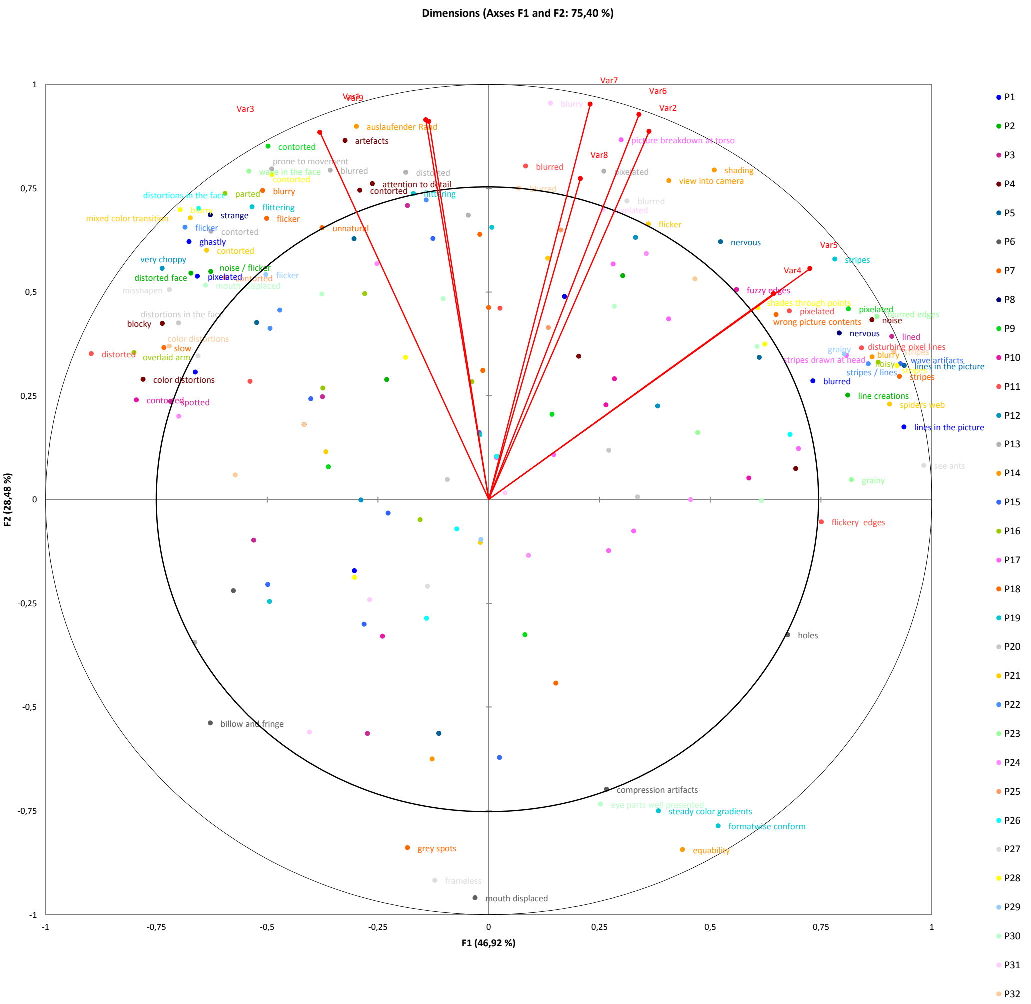


Figure 4.19: Perceptual space and perceptual directions (red lines) of TR A4-S5.

Interpreting Figure 4.20 by considering the perceptual directions leads to the assumption that attributes such as pixelations, blur, and flicker are perceived as the most severe ones.

For A5-S5, the first two components (F1 and F2) account for 74.10 % of the explained variance (dimension 1, F1: 55.79 %, dimension 2, F2: 18.31 %). These two dimensions, as presented in Figure 4.21, form the perceptual space. Attributes with a R^2 between 100 % and 50 % are holey, incomplete, blurry, flicker, pixelated, edges, wrong contours, flicker at contours, nervous, noise, contorted, lines, fringed edges, deformations, and grainy. TR A5-S5 seems to have the same problems as the previous TRs; however,

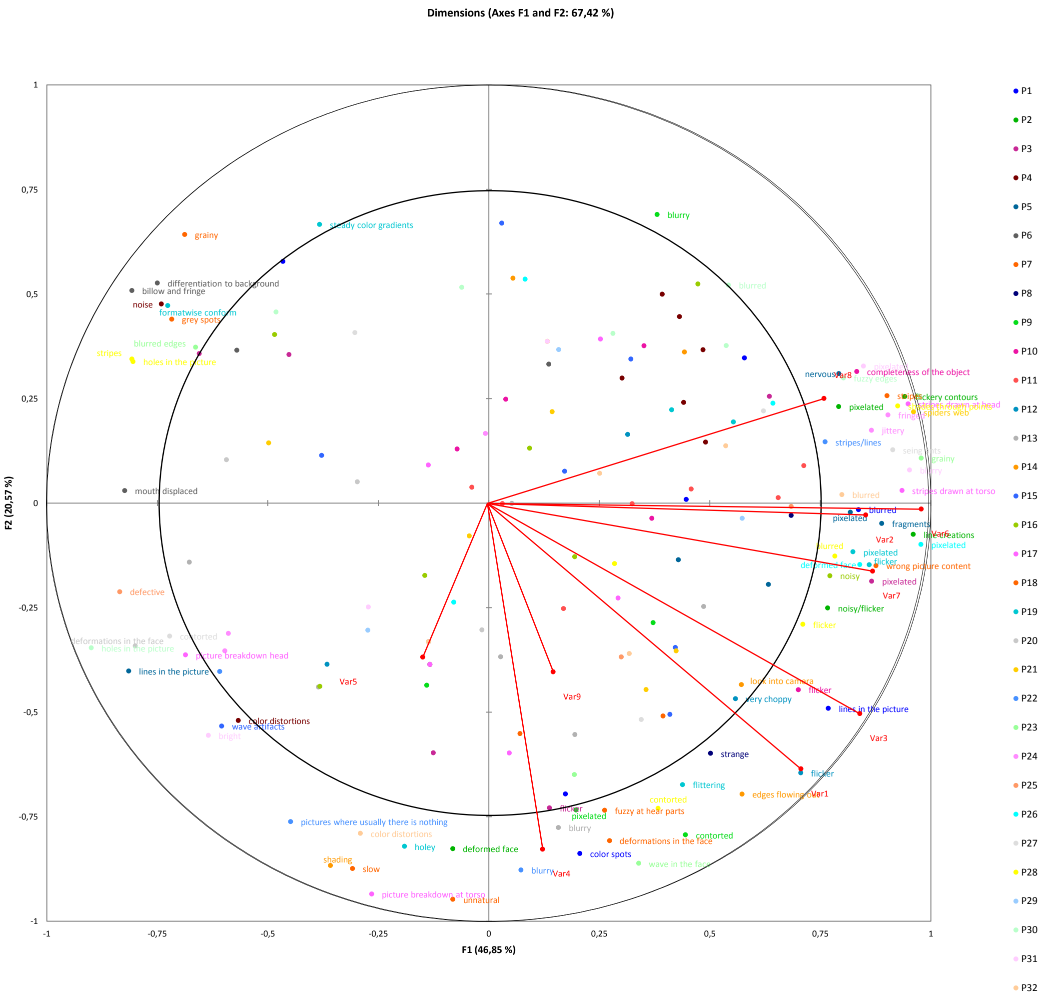


Figure 4.20: Perceptual space and perceptual directions (red lines) of TR A5-S4.

fringed edges, flicker at contours, and holes are mentioned here more often than for the others.

The perceptual directions (red lines) of Figure 4.21 lead towards the directions of flicker, pixelations, lines and stripes and, therefore, these attributes seem to be relatively severe.

To sum up the differences between the five presented perceptual spaces, one can interpret that, at first sight, all of the used TRs seem to have similar quality describing attributes and exhibit only slight differences with varying degrees of severity. Here, the question was how strongly the various TRs differed. Therefore, a further analysis and interpretation

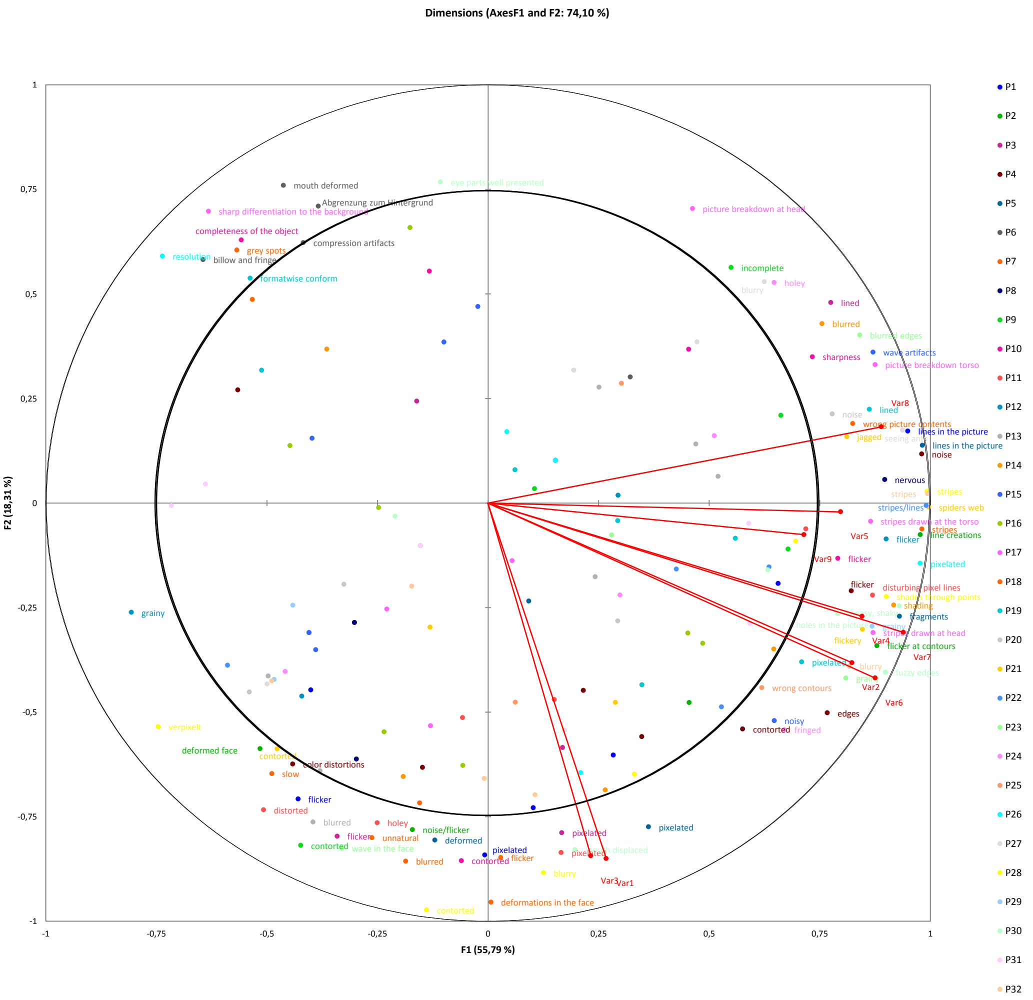


Figure 4.21: Perceptual space and perceptual directions (red lines) of TR A5-S5.

concerning the attributes' direction within the correlation circle and the definition of confidence areas for significance elicitation, as described in [130], was carried out in a next step (i.e., perceptual directions). Further differentiations could be represented in the quality taxonomy that is presented in Section 5.3.

Results of Eye Contact Perception

The measured eye contact perception shows the indicated differences in perceived or not perceived eye contact by the participants (Chi-square test, $p < .0001$, df = 43). These results are from the investigation of all the available data derived from the rating of all 44 test items.

Obvious differentiations are made between the man presented without eye contact (3.13 % perceived eye contact), as shown in the third column of Figure 4.10 in Section 4.2.3, and the other three content variants (59.5 % perceived eye contact for the man in the first column, 85.8 % for the man in the second column, and 86.25 % for the woman displayed in the fourth column of Figure 4.10).

Taking into account the different creation processes across all contents and excluding the ratings of the man without eye contact, we see that there are no significant differences (Cochran's Q, critical: 9.49; observed: 8.95; $p < 0.062$; $\alpha = 0.05$). Analyzing differences of the eye contact perception based on the used TRs per presented content, significant differences can only be found for the content showing the man without eye contact (Cochran's Q, critical: 18.31; observed: 272.65; $df = 10$; $p < 0.0001$; $\alpha = 0.05$). Here, the distinctions are made once between TR A5-S4-BG and all other variants, and once between A5-S5-BG and the other TR variants.

Results of Questionnaire on Media Usage and FPI
Data about demography were already presented in the description of the test participants in Section 4.2.3. It was noted that the participants were naïve, which means that they were participating in such an evaluation process for the first time and did not have further knowledge about the creation process of the tested stimuli or had seen such test items previously. In the questionnaire, the self-estimate for media experience was made on a 7-point scale, between 1 = very familiar and 7 = not familiar at all. The average media experience turned out to be 2.31. Asking about the usage of different media applications in particular (between 1 = several times per day, 2 = daily, 3 = several times per week, 4 = weekly, 5 = monthly, 6 = more seldom, and 7 = never) makes it possible to draw a more detailed picture of the participants: The majority (30) uses the Internet several times a day and two use it daily. A cell phone is used daily by the majority (14) and ten use it several times a day. Radio is used by everybody and several times per week by the majority (9). Only six participants use it several times per day. TV consumption differs, because the majority (13) watches TV every day, and only four use it monthly. 3DTV is used never by the majority (25) and seldom by seven participants. The PC is used several times per day by the majority (23). Video telephony is used monthly by the majority (12), seldom by nine and never by six participants. Only two participants use video telephony several times per day. The answers to the FPI show that the majority of the participants is rather sociable and extroverted. There was no data set with a contradictory result.

Summary of Presented Results

Concluding and comparing the results collected in this study, leads to the conclusion that rather communicative participants who are relatively familiar with media perceived the quality of the presented TRs as not acceptable, but rated the various TRs differently. Comparing the results of the psycho-perceptual ratings with the sensory data supports the previous assumptions that holes and missing picture parts are perceived as most severe. TR A2-S3 was rated worst compared to the other TRs, and the quality describing vocabulary includes attributes related to the aspects holes and missing picture parts more frequently than the other TRs. The TRs A5-S4 and A4-S5, the second worst rated, are not clearly differentiated. Their psycho-perceptual results are supported by the quality describing attributes, which are pixelations, blur, and flicker, for both TRs. However, for TR A5-S4, the quality describing attribute holey (i.e., incomplete image having holes) is also mentioned in another way than TR A4-S5. The TR A5-S5, is described with quality attributes such as flicker, pixelations, lines, and stripes. The TR A4-S4, rated less bad, was described by noise, blur, fragments, pixelations and unnatural representations, and with picture breakdowns at the head and face parts. However, among these describing attributes, a dominant quality attribute, compared to TR A2-S3 with the aspect holes, does not seem to be present. The quality differentiation given by the results of the psycho-perceptual ratings and supported by the sensory profiles is, similarly, represented by the eye contact rating. Here, TR A2-S3 is rated as the TR for which eye contact perception is rated less, TR A4-S5, A5-S4 and TR A4-S4 are rated similarly. For TR A5-S5, eye contact perception is rated less often than for the other three (i.e., A4-S5, A5-S4, A4-S4), but not as seldom as for TR A2-S3. Nevertheless, as mentioned in the results of eye contact perception, no significant differences were detected and the influence of background presentations on the ratings has to be investigated further. In a next step, a comparison of the descriptions of the TRs and the problems considered in this context with the presented evaluation results is regarded as useful for the extended definition of a quality taxonomy in Section 5.3.

Taking the method presented by Lorho [151] for the interpretation of direction and the method of confidence ellipse interpretation, presented by Husson et al. [202], into account, leads to the requirement for an add-on PCA. For the interpretation of direction, this PCA could be performed on the group average data provided by the GPA conducted previously. For the confidence ellipse interpretation, the raw data used for the GPA could also be used for the PCA. The advantage of these interpretation methods would be that they provide additional information in the plots for further interpretation and support for significance estimation by the confidence ellipses, if needed.

4.3 Summary - Applied Evaluation Design

The research question was: which quality features (QFs), perception-based quality influencing factors, can be detected in order to describe the quality of the TR used to create 3DVO for the described use case video communication. In a next step, the question becomes: which aspects of this knowledge about QFs are useful to profit from the TR process and are useful for the quality taxonomy in particular. Therefore, the evaluation was conducted for the used TR algorithms and the perception of produced quality. This section outlines and discusses the results of the applied quality assessments in detail. More or less severe influences of QEs and information about the interaction between different quality influencing factors can be registered on the basis of the findings on QFs as a result of the evaluation activities.

All in all, results of a literature analysis, three expert interviews, two focus groups with eight participants each, an online questionnaire answered by 25 participants, a laboratory experiment with more than 300 participants, and a further laboratory experiment using OPQ with 32 participants, are presented in detail in the related sections. From these results, a list of detected QFs and a collection of useful knowledge is summarized for the proposed concept of linking TR with QoE. This takes into account knowledge about perceptual concerns in the results derived from user-centered evaluation approaches. However, the reason(s) why something happens (e.g., within (re)cognition) is often unknown. For the presentation of the results, the following common steps for data analysis have been considered (after Zacharov [90]): (1) check the data and variability, (2) check the assumptions, (3) panel performance, (4) use the methods (for data analysis), (5) test the hypothesis (question) - null hypothesis, (6) conclude. Based on this practice, and based on the various methods for data analysis presented in Section 3.3.6 and Section 4.2, results have been presented and their validity and reliability, as derived from these methods, are discussed in the next section. The focus of this discussion is on their usefulness for the intended quality taxonomy definition and the possibility to use them in the proposed approach for the extended quality taxonomy in Chapter 5. All in all, an explorative and systematic approach was applied in order to define QFs defining the QoE related to the presented TR processes created within the project Skalalgo3d.

4.3.1 Relation to the Area of Application

Results of the subjective assessment of developments aimed at eye contact support in video communication systems via 3DVO usage and useful for the intended interlinking model have been presented in the previous sections. It is intended to realize the linkage of quality assessment of 3DVO with algorithm development, with support of defined quality influencing factors, their weighting, thresholds, and specific determinants. In order to define perceptual thresholds and severity levels of detected QFs, further analysis with consideration of real GT data and applications that are real-time capable may lead to more detailed results, including the influence of interactivity, which is considered to be a relatively severe influence. The following results from the methods described and implemented above are used for the interlinking model:

- EXPLORATORY ANALYSIS: Qualitative description of presumed quality influencing factors and quality describing attributes (lead to the definition of QFs (and QEs)) as well as presumed influences on acceptance or non-acceptance of perceived and produced quality. For the latter it is assumed to lead to the definition of links.
- LABORATORY EXPERIMENT: Acceptance rating (yes/no, gives an overview of quality), defined quality influencing factors based on qualitative description of perceived quality and their frequency which lead to definition of QFs, and presumed quality influencing factor eye contact (perceived/not perceived) which may lead to the definition of QFs, but, because it was elicited explicitly for eye contact per caption, it is not used as a user-centered QF definition and therefore not further considered.
- OPQ: Acceptance Rating (yes/no, gives an overview of quality), ACR of quality describing attributes defined by users which lead to the definition of QFs), eye contact perception rating (yes/no) which may lead to the definition of QFs, but, because it was elicited explicitly for eye contact perception, it is not used as a user-centered QF definition and therefore not further considered.

This information results in the definition of QFs and influences exerted on them by particular QEs, which form the particular TR. With this focus on specified results useful for an interlinking model, the analysis is conducted and results are described explicitly in Chapter 5. In doing so, mainly frequency information is used to investigate the definition of QFs with severe influence on quality.

4.3.2 Applicability of Results for Interlinking QE and QF

Defined QFs are used and the influences on them by particular QEs that are present in the particular TRs. With this basis, the focus of the next steps lies on specified analyzed results useful for the interlinking model to be presented in Chapter 5. By comparing the results of all experiments, differences, similarities, recognizable aspects (summary of QFs and first relations to QEs) are considered, if they are obvious from the results. The most useful information for an interlinking model derived from the conducted evaluation activities are the frequencies of the mentioned QFs representing their severity. They are interesting, because, to date, no definition of important QFs has been considered that evolves from a TR as presented within Skalalgo3d. The results show that the detected and investigated QFs clearly differ from the common, well-known problems such as resolution. However, naturalness and sharpness are a remaining problem. The information about further detected QFs is useful to define a more global quality taxonomy, leading to a quality measure that includes more user-centered information; in addition to the quality influences by the TR, and in order to define interrelationships and dependencies. In future investigations this may support the adaptability of the TRs to certain circumstances and scalability related to available resources but still leading to an acceptable quality perception on the users' side.

5 Linking QE and QF: Defining an Extended Quality Taxonomy Model

The objective measure 3DVQM for the quality assessment of 3DVO, as introduced by Rittermann [26], was presented in Section 3.3.2. There, the minimal consideration of the subjective quality assessment was criticized. Hence, a more user-centered approach is intended. In order to provide a quality measure that represents the overall quality, QoE should be considered as comprehensively as QoS. The author considers overall quality in this thesis as a composition of physical and perceived quality, defined by the QoS and its QEs, on the one side, and QoE and comprehensive QFs on the other side, as well as by their interchangeability (see Section 3.1). This leads to the research question of this thesis (as outlined in Section 1.1) *how the technical realization process of 3DVO creation can be interlinked with the definition of resulting quality and, in this way, profit from subjective quality assessment.* The supplementary questions were *how* to interlink *specific* QFs and QEs. The QEs are defined by the use case and its TR processes, as described in Section 2.3.2). The QFs (besides overall quality via ACR) have been defined by a quality assessment that asks *why* quality is rated as good or bad (see Section 4.2). On this basis, an extended (taxonomy) model demonstrating the knowledge about quality influencing factors is presented in this Chapter 5. This includes information on how this knowledge can benefit prospective technical development and processing algorithms of free viewpoint video objects. Therefore, the interlinking model development is considered, in general, in a first step. In a second step, different approaches relevant for the interlinking idea are presented. In a third step, the choice of the extended taxonomy model is presented by outlining its strengths and weaknesses and providing an exemplary proof of concept.

5.1 Interlinking Quality Taxonomy Model

The basic idea of an extended taxonomy model is to

- give an overview of quality influencing factors, their potential relationships and
- provide a concept to use this knowledge by quantifying the relations between QFs and the underlying QEs.

Roufs and Bouma [98] already presented an approach to linking perception research and image quality in 1980. There, questions on the factors of image quality and the use of thresholds as limits of perception, as well as considerations of modeling itself, were addressed. The concept of modeling, as described by Ludewig and Lichter (see [89], pp. 5-8), is used to represent the actual state and the prescriptive state of a certain condition of interest in this case, the relationships between TR and QoE of 3DVO. This means that iteration processes are considered and the model may change over time, based on further knowledge and/or changes within the TR process or the QoE.

The proposed interlinking model intends to make use of knowledge about QF and to benefit the TR of 3DVO for eye-contact support in video communication systems. The intention is to make use of a theoretical, clearly structured, flexible, adaptive, and scalable model for algorithm development. Therefore, several requirements have to be considered in order to define a useful and usable model. The model should:

- use information about the considered TR and accompanying QE, as its basis,
- take into account changes of the considered TR and, thus, changes of the QE,
- allow the presentation and the adequate usage of the knowledge about QF (derived from previous subjective quality assessment, see Section 5.4),
- consider influences from QE on QF and vice versa, and
- reliably display the links between QE and QF.

The interlinking model has been built on the basis of these requirements and on the concept of modeling presented by Ludewig and Lichter [89] and Roufs and Bouma [98] who suggest to represent givens, possible changes, and relationships. In order to define the most adequate approach, various concepts have been considered in a systematic literature analysis (see Kepplinger [18] and Section 4.2.1 part (B)). These concepts are presented briefly in the next section, to provide an overview of existing possibilities and to outline the reasons why the taxonomy approach is used at this stage of the work.

5.2 Approaches to Interlinking QoE with QoS

In the introduction in Section 1.2.2, Silzle's [4] approach to interlinking QoE with the TR processes (and QoS) has been presented, besides others. Focusing on representative dependencies, the levels of influences and an adequate weighting, metric-based concepts, and their pros and cons will be considered in this section. The viewpoint in this work is empirical and driven by activities within HCI and QoE-related research. For this reason, it will not include further approaches provided in machine learning and computer science communities or the use of machine-learning-based approaches for prediction (e.g., as presented by Shahid et al. [204]). However, it might be advantageous to consider these approaches as a next step of the present activity to interlink the knowledge of QoE and QE once quality influencing factors and their interrelations are defined.

Several concepts found in a preliminary literature review (see description of literature analysis (B) in Section 4.2.1) are outlined here to present the basic approaches considered. This is done to locate the most ideal approach for an interlinking model. In addition to the approaches already presented in Section 1.2.2, different ideas and models to integrate knowledge about QF or evaluation results on QoE into the creation process (TR) are presented. These include the approach to instrumental metrics presented by Reichl et al. [205] that has arisen in the QoE community, or the approach introduced by Skorin-Kapov and Varela [14] and mentioned in Section 3.1. This is, namely, the ARCU model, which focuses on the categorization of quality influencing factors into the spaces application, resource, context and user, and which currently is under further development. Following the presentation of a collection of the advantages and disadvantages of each concept, an adequate concept for an interlinking model in the context of this work is finally discussed. Some comparable attributes to differentiate the considered approaches are, for example, traceability, semantic gap, elaboration, efficiency, calculability, and usefulness. These comparable attributes will play an important role when it comes to the definition of a useful concept for a model that interlinks QoE with algorithm development. Towards this, in the following concepts like norms, recommendations, design patterns, heuristics, semiotics, ontology, taxonomies and QoE Models found in the mentioned literature review are presented in short:

5.2.1 Norms

The advantages of norms or normative concepts are that they enable reasoning about actions and beliefs in a flexible and logical way without being a strict law or rule (see Douglas et al. [192]). This means, that it does not have regulative or monetary consequences as legal effect, if the norm is not considered. This provides measurement and a basis for comparison of certain states. Norms help to avoid disputes and to aggregate complex work flows or processes in an efficient and calculable way. In general, norms may help at an early stage of defining the relationships between QE and QF and provide an overview of the user behaviors, QoE, and QF.

5.2.2 Recommendations

According to the dictionary [206], to recommend something means to "put forward (someone or something) with approval as being suitable for a particular purpose or role". This understanding is similar to norms. However, recommendations are usually understood as choice assistance. (The term recommendations should not be confused with Recommendations (capitalized) provided by ITU which are standards.) Some examples are recommendations given to the user e.g., suggestions about which content to watch on TV, based on known preferences, or recommendations supporting design or evaluation processes. Thus, there are differences between recommendations for a *must have* and those denoting a *nice to have*. The advantages of recommendations are that they provided guidelines, allowing an orientation and prioritization. Similar to norms, recommendations do not exclude the freedom to decide. One disadvantage of recommendations is that they are inconclusive. Furthermore, it is time consuming to set up recommendations, especially when they refer to *must haves* and need to undergo a validation. Additionally, the representation of the real world is questionable, especially in contexts where things are changing rapidly over time or where it is impossible to construct a complete representation of the setting.

5.2.3 Design Patterns

Design patterns are defined as a documenting architecture of the intended design in the form of an essay that contains information about the aim, motivation, structure, behavior, applicability, and consequences (see, e.g., Gamma et al. [3]). Sometimes, relationships to other patterns are included. They provide support in solving design issues in a concretized and transparent way. The advantage is the specification of problems as well as the description of its possible solutions. However, design patterns include a high risk of misunderstanding and, hence, misuse, which may not be useful in a complex, dynamic, but, at the same time, deterministic (software design) environment (see also Hsueh et al. [170]) as is the case in the presented use case.

5.2.4 Heuristics

Heuristics in computer science help to aid and control problem-solving in a time-saving and in an approximate way. Their intention is to find a solution for the problem at hand very quickly without being necessarily very accurate, precise, or complete. This requires a decision about whether the solution should be the best solution available or a fast solution (see also Cheng et al. [193]). Nielsen [207] points out that heuristics are employed in usability and quality evaluation work, as a basis for a systematic inspection, and to locate usability problems in a system or application. Again, rapidity plays an important role as a heuristic evaluation is a fast and easy way to assess, for example, user interfaces and to obtain rapid feedback on severe problems, without a claim for completeness. Heuristics, for the purpose of interlinking QF and QE, may be useful to detect influencing issues and coherences in a real-time application without claim to completeness, but not for the criteria to work as an interlinking factor on its own—at least without recurrence.

5.2.5 Semiotics

Semiotics are used in the HCI field in order to explain the designer-to-user communication. This includes the classification of relationships and associations (e.g., between content and expressions) into signs. According to Bento et al. [177], these signs can be static (no temporal or causal relations), dynamic (depend on temporal and causal relations), or metalinguistic (refer to other signs and are based on dependencies). Semiotics have a long history and are used in various areas of science. For example in sociology, medical science, communication science, computer science, etc. Semiotics work as a communication basis and are usually divided into several branches. The advantage is that they provide coherency and support traceability as mentioned in [18].

5.2.6 Ontology

Ontology is a basic philosophical discipline that asks about the basic structure of entities and reality. It also has a long history. It not only investigates what it means to be, to exist, or to be a positive something but also investigates various classifications of existing things (e.g., alive/not alive, physical/not physical, concrete/abstract, individual/general). It theoretically formulates and compares the ontology of different models on the ontological assumptions of our language and describes or replaces them with better fitting assumptions (see Brockhaus Encyclopedia [208]). This concept might be useful for mathematical formulae and calculations of relationships and influences through the use of different forms of categories, classifications, and weightings as described in [18].

5.2.7 Taxonomies

The word *taxonomy*, as mentioned in the German Duden Spelling Dictionary [121] is derived from the Ancient Greek *taxis*, which means “arrangement”, and from the Ancient Greek *nomia*, which can be translated as “law” or “method”. The simplest meaning of taxonomy is: “to regiment into a system”. Common areas of application for taxonomies are, on the one hand, botany and zoology and, on the other hand, linguistics. The first-mentioned areas employ classification schemes for the placement of living beings into categories. The last-named area aims at describing the language system by segmentation and classification of linguistic units [121]. However, the areas of application are even more varied. For example, the concept of taxonomy, as used by Silzle [4], was already mentioned by Moller [209] in 1977. The object of study was different, but the domains are defined similarly. The advantage of taxonomies is their traceability. Amongst others, according to Silzle, “a taxonomy is a special case of an ontology” (see Silzle [4], p. 56). As already introduced by Silzle [4] or Möller et al. [171], taxonomies work as a breakdown of physical elements of an investigated system and provide an overview of identified perceptual factors of the user. They visualize identified relationships and influences on the overall quality, e.g., by taking into consideration factors of QoS and QoE. This is especially useful when considering overall quality as a weighted sum of QoS and QoE, as done by the author in this thesis.

5.2.8 QoE Models

Generic approaches and microeconomic models that take QoS and service dependent influence factors into account are introduced in Section 3.2. Their advantages are the use of a comprehensive and holistic approach to investigate contextual and user-centered factors, as well as system-based factors, including mutual correlations and weighted combinations going beyond simple MOS values for the definition of QoE. The logarithmic laws introduced by Reichl et al. [205] are promising. They will define instrumental metrics, to “[...] provide sustainable mappings between traditional QoS parameters (which are strongly technology focused and thus easy to measure (sic!)) and efficient user-centric QoE metrics”. This approach puts QoE evaluation results about user experience and satisfaction, which follow logarithmic laws, into the context of the Weber-Fechner Law, a key concept in psychophysics. This law describes the relationship between a physical stimulus and perception (see also [205]). However, much information about different logarithmic relationships is lacking, especially when it comes to more complex relationships within complex systems and indicators. Examples for this are the various TR algorithms (e.g., A1-S1 vs. A2-S2 as introduced in Chapter 4) in correlation to QEs with other than simple relations (e.g., loss rates in correlation to MOS). Notwithstanding the above, the approaches dividing influencing factors into classes seem to be very useful, if not even unavoidable.

5.2.9 Summary and Comparison of the Presented Approaches

Taken together, norms, recommendations, and heuristics have similar constraints, based on their broad rules, which, at first view, do not seem useful for implementation in the presented use case. Semiotics are becoming increasingly popular within the field of HCI and artificial intelligence. As already mentioned, a few implementable approaches have been described. The majority of them suggests further development using ontology. The QoE models are confirmed with simple subjective assessment metrics or pre-defined influencing factors, also suggesting the further development of metrics. However, these may function for entities that have already been defined and validated. This is not yet the case in most of the presented use cases. The approach of creating a taxonomy as presented in [4] meets the requirement to use information about the considered TR and accompanying QE as a basis. Therefore, this approach is seen as being useful for intended purposes, after a modification related to user inclusion, in order to meet the requirement of a user-centered taxonomy (see next section).

5.3 Proposed Concept

In the following, a taxonomy approach that leads to an interlinking model and a structure chart is presented. This is based on theoretically outlined quality influencing factors presented within the description of the model's physical domain and its QEs (see also Chapter 4 and Weigel [2]), as well as within the perceptual domain and its identified QFs, based on results of the subjective quality assessment (as described in Chapter 4). The differences to the other concepts that have been previously described in this work are its requirement to function as an visual overview for further development (e.g., algorithm development) without being, as yet, restrictive and formalized. The advantages are the overview that it provides and that it can be used as a tool to create patterns or rules for adaptive or adjustable and scalable algorithm development in the future. The disadvantage is that it is a graphical overview, in a first step, that does not provide a mathematical description of interrelations, dependencies, influences, and relationships, although they are organized in terms of their severity and degree of influence. However, with the knowledge about the importance and severity of certain QFs and with the usage of a further developed TR (i.e., real-time capable application) and the combination of collected knowledge with QoE parametrization tools (e.g., eye tracking combined with biofeedback recording systems) the lack of reliable, mathematically defined parameters may be overcome in a future step. Different ways to parametrize quality influencing factors are already described in Kayser [112] and Fechter [94] (see also Rittermann [26]), or Baumeister [210], as well as Wolf and Pinson [139]. There, factor analysis is the dominant method. Focusing on representative dependencies, levels of influences, and an adequate weighting, the metric-based concepts and their pros and cons will receive more attention in the next sections. Current activities aimed towards parametrized data usage in the video context include further parametrization tools such as eye tracking, for example for a saliency-based video quality prediction, as presented by Redl et al.[211].

In a first step, the results of Study 2 (laboratory experiment) and Study 3 (OPQ) are mapped into a quality taxonomy, similar to the procedure introduced by Silzle [4], but not by asking experts to map the QFs to the respective QEs. Instead, the findings of the conducted user-centered studies are used. Therefore, the detected quality influencing factors of Study 1 (explorative evaluation) are used to define relevant QFs by a qualitative analysis using cluster analysis, as proposed by Glaser and Strauss [161] within the grounded theory approach. These are complemented by the results of Study 2 and Study 3. This occured in each case through clustering of the mentioned QFs according to the test participants' quality descriptions (see Section 4.2). This resulted in the definition of 12 relevant QFs. These are:

blurred
color distortions, bad color quality
color spots, blemished
contorted
deformations, shape distortions
flicker, jitter
incomplete, holey
pixelated
slow lagged
stripes and lines
unlikely background
wrong proportions

An explanation of these 12 QFs is given in the related sections of Study 2 and Study 3 in Chapter 4 and a translation is provided in the Appendix A. For the mapping of the QFs perceived by the user the results of Study 2 and Study 3 are used. Therefore, the frequency of the most important attributes as defined via GPA and presented in Section 4.2.3 (i.e., attributes under consideration of the determination coefficient of correlation circles, the perceptual spaces) are considered for each evaluated TR procedure (i.e., different combination of creation steps, the QEs).

The results are presented as scatter plots in Figures 5.1 and 5.2. The size of the bubbles indicates the assumed importance of the respective QF, defined by the number of indications (Study 2) as well as the number and severity of mentions (Study 3). Here, for example, in Study 3, the test participants had to define their quality describing attributes and the severity of the attributes' presence. Based on this, a weighted severity factor over the test participant's ratings for each QF was defined (see Study 3). The different QEs are described in Section 2.3.2 (see Table 2.3 on stereo analysis and Table 2.4 on synthesis) and used for the test stimuli in Study 2 and Study 3. As presented

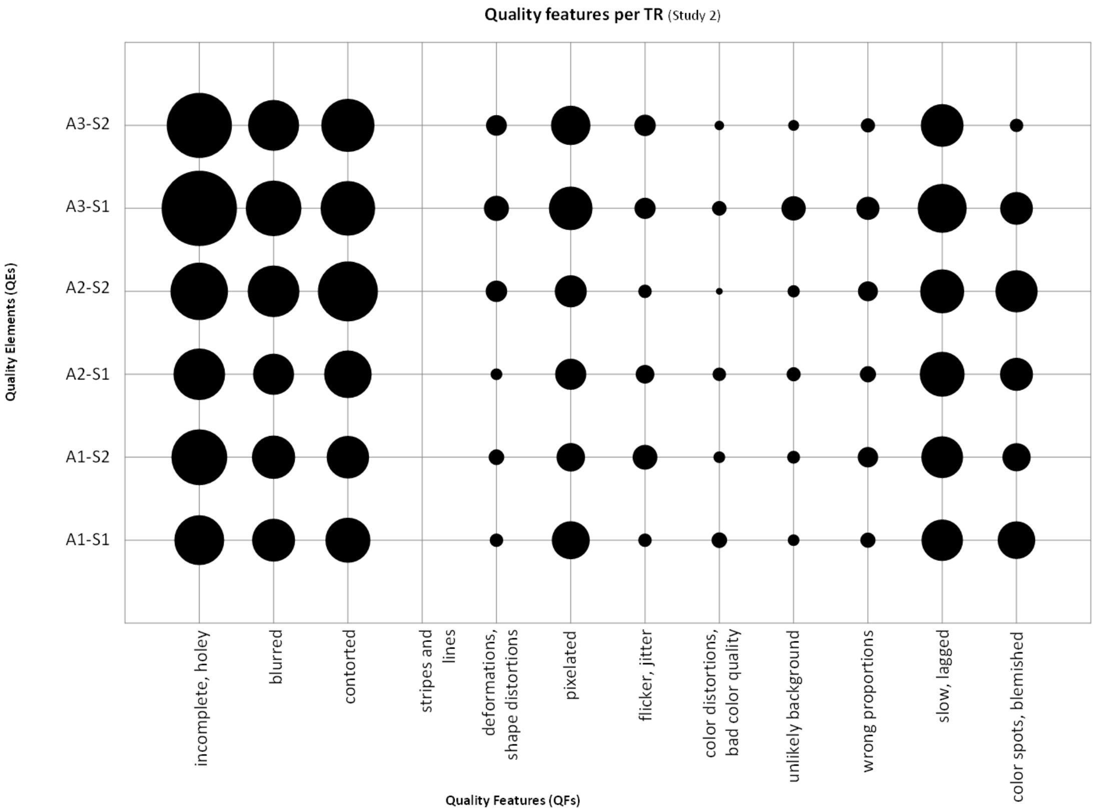


Figure 5.1: Quality taxonomy derived from Study 2.

in Figure 5.1, in a comparison of Study 2 with Study 3, the QFs stripes and lines are not mentioned often. The most common QF is clearly incomplete, holey. This is accompanied by contorted and blurred. Pixelated and color spots are more present than deformations/shape distortions, flicker/jitter, unlikely background, wrong proportions, and color distortions. Slow, lagged was a problem in Study 2 and is referable to the equipment available to present the test stimuli, as described in Section 4.2.2. This perceived QF has clearly to be considered as a quality factor for interaction in real-time capable video communication systems, as discussed in [23].

Comparing the QEs used in Study 2, it seems that there were rather small differences in quality perception related to the QFs (see Table 6.4). The assumption is that the QF incomplete/holey was so severe that it led to the disregarding of further problems, besides contortions and blur in contrast to the results of Study 3. These latter two QFs also seem to play a very important role. For example, for A2-S2, contortions (234 counts) seem to be slightly more severe than incomplete/holey (215 counts). Compared to the results presented in Figure 5.1, Figure 5.2 shows more differentiated results when comparing the different QEs. The QE A2-S3 is the only one that has one part of the TR in common with the TRs used in Study 2, namely the way of stereo analysis (A2). The scatter plot in Figure 5.2 shows a clear differentiation to the other QEs, especially for the QFs incomplete/holey and contorted.

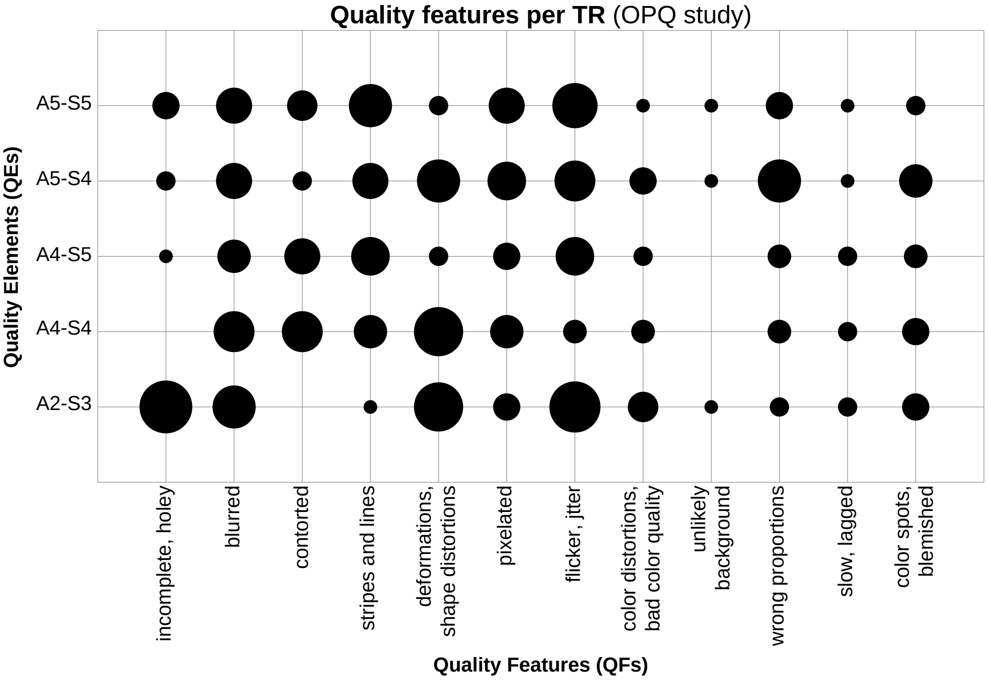


Figure 5.2: Quality taxonomy derived from Study 3.

These taxonomies fairly represent the results presented in Section 4.2.1 and Section 4.2.1. The quality taxonomies provide an overview of quality influencing factors and their potential relationships and provide a concept to make use of this knowledge by quantifying the relations between QFs and the underlying QEs. For instance, this is possible with usage of the size differences of the bubbles (i.e., significance tests of frequencies, etc.).

As seen in Figure 5.1, the elements incomplete, holey, contorted, slow lagged, as well as blurred and pixelated seemed to be more severe than others. For example, color spots and blemished are a little weaker and stripes and lines are least severe. Comparing these results to the results of the second quality taxonomy shown in Figure 5.2 and to the underlying TRs leads to the assumption that there are more and less severe quality defining factors.

So far, no weighting factors, confidence intervals, or levels of variance are shown within the presented quality taxonomies, because these taxonomies are based on frequency values created from the results previously presented in Chapter 4 (see also Appendix B). However, the size of the bubbles, as presented in Figures 5.1 and 5.2 is used as a severity indicator. Additionally, a proof is intended with the conduct of a further analysis using the Analytic Hierarchy Process (AHP), as introduced by Saaty [212]. This is done by conducting an expert evaluation, which is presented in the next Section 5.3.1, in order to provide a Proof of Concept (PoC). A comparison of levels of severity indicated by the extended quality taxonomy with accompanying bubble size and by the expert evaluation is carried out. Based on this, an exemplary suggestion about how to include evaluation results (weights) into the introduced 3DVQM is given as a final step of this thesis.

5.3.1 Proof of Concept

A PoC for the investigated application at this current stage is conducted by considering an AHP by an expert and an example of how to use the resulting quality taxonomies in a formalized measure, the *Subjective* 3D Video Object Quality Metric s3DVQM. For this purpose, further threshold definition is seen as being useful in a next step in order to define boarders of interrelationships or perceptual boarders and make usage out of them for scale able and adaptive applications. Weighting of by subjective evaluation defined QFs to further represent differences of severity and influence on overall quality beyond test participants' perception by including the QEs influences indirectly is seen as being useful for now. The reason therefore is, that technical measurements, e.g., amount of perforation in percent, or similar values, are not available thus far. In a further step, a more detailed investigation of single QEs and a sublimation of one QF to another, in order to isolate influencing factors from each other and to further define their interrelationships, perception thresholds, and influencing severity, has to be targeted. However, for now, a validation of the quality taxonomy presented above and its inclusion into the TR process is given in the subsequent expert evaluation.

5.3.2 Model Refinement by Considering an Expert Opinion

In order to consider different severity of quality influencing factors from both sides, the TR and the subjective perception, a weighting of QFs by the executive producer (the one who realized the investigated test stimuli) is intended. Therefore, the pre-defined 12 QFs presented in the evaluation results in Chapter 4 (see enumeration at the beginning of this section) are ranked to each other by a pairwise comparison. This occurs on a scale of 17 steps (from 9–1 until 1–9), indicating to which extent one QF is more important than the other one. This is conducted as suggested in [212]. The results are used in order to compare the opinion of the expert with the results derived from evaluations with non-experts (see above).

One expert (male, 33 years old, familiar with the 3DVO creation process) answered, for 132 pair combinations using the scale of 17 steps introduced above, the following question: “How many times more important is one criterion (i) than the other (j) with respect to the overall quality of the evaluated first intended representations?” The first intended representations are the used test stimuli, see also Figures 4.2 and 4.10. The expert performed the TR for the test stimuli shown in the previous user-centered evaluations, but he did not watch the 3DVOs explicitly for the AHP. Saaty [212] points out that the answers are consistent when

$$A_{ij}A_{jk} = A_{ik} \quad \forall i, j, k \tag{5.1}$$

As described in [4], A_{ij} on the left is the relationship between criterion i and criterion j based on the answers. Because the method assumes reciprocal answers ($1/x$) from the test participant, the lower diagonal part of the matrix A (of order n) is not considered further. In order to check the consistency of the answers given by the participant(s) of the AHP, the matrices from the answers and their eigenvalues are investigated. In this case, 66 pair combinations of the 12 QFs remain because they are the upper diagonal part of the matrix A. The consistency ratio (c_r) of an individual participant has a value of < 0.1 (see Equations (5.2) and Equations (5.3)). λ_{max} represents the principal (maximum) eigenvalue of the matrix A. Usually, c_r of a result matrix A should be smaller than 0.1, because 10 % inconsistency is tolerated. This means that the participant of this exemplary AHP has answerd in a consistent way.

$$\mu = \frac{\lambda_{max} - n}{n - 1} \tag{5.2}$$

$$c_r = \frac{\mu}{\mu_r(n)} \tag{5.3}$$

The random consistency index μ_r is provided by [212], where the definition of these values is explained by using a randomly generated reciprocal matrix from a sample of size 500.

Because results gathered only from one participant are considered, neither the consistency of several judgments of experts is analyzed, nor a mean and 95 % confidence interval calculated. However, this should be considered when results of an AHP with more participants are provided. An example of how to calculate the consistency ratio and how this ratio is given is described in [212] (as well as in [4], p. 69).

Considering the principal eigenvector of the matrix created by the answers leads to the weighting factors. The ranking values are presented in Table 5.1 (see severity coefficients) and emphasize that all QFs are important, but incomplete, holey has a higher value, compared to the other attributes, and seems to be the QF with the most influencing weighting factor. This is in conformity with the previously reported results that *incomplete, holey* is a relatively dominant QF.

Comparing these results to the extended quality taxonomies presented above (see Figures 5.1 and 5.2) and to the results presented in Sections 4.2.1 and 4.2.1 leads to similar results concerning the QFs' magnitude of influence. Nevertheless, in contrast to these evaluation results, the expert rated blurred and, similarly stripes and lines as being not very severe. The attribute incomplete, holey is rated as the most severe by the expert, followed by contorted and deformations, shape distortions with equal severity, and wrong proportions. In contrast, the expert found the QF wrong proportions more important, whilst it was not mentioned explicitly by the test participants of Study 2 and Study 3. Based on this, the following *severity coefficients* can be defined for the QFs (see Table 5.1): These values are used as initialization values (i.e., coefficients (β_i)) for exemplary usage of the 3DVQM, based on available results, in the following.

Table 5.1: Severity coefficients derived from an AHP with one expert.

QFs	severity coefficients
blurred	0.02
color distortions, bad color quality	0.03
color spots, blemished	0.06
contorted	0.16
deformations, shape distortions	0.16
flicker, jitter	0.05
incomplete, holey	0.20
pixelated	0.05
slow, lagged	0.03
stripes and lines	0.07
unlikely background	0.03
wrong proportions	0.13

5.3.3 Inclusion of Results (Weights) into 3DVQM

As described in Section 3.3.2, the 3DVQM, as introduced by Rittermann [26], uses a combination (weighted sum) of parameters ($p_{\psi,i}$) that depends on the chosen coefficients (β_i):

$$\mathbf{3DVQM} \equiv \sum_{i=1}^{N} \beta_i \cdot p_{\psi,i} \tag{5.4}$$

Therefore, a change concerning the quality parameter $p_{\psi,i}$ (e.g., either they are modified or enhanced) also leads to a change of the related coefficients as weighting or limiting factors. Because the measure is used as described in [26], but the weighting coefficients are based on the subjective assessment of quality instead of a theoretical mathematical derivation from a TR point of view, the measure is named s3DVQM in the following. Using the previously defined 12 QFs and their values presented within the quality taxonomies as the quality parameter $p_{\psi,i}$ instead of TR-related, theoretically defined factors only, and previously defined weighting factors as coefficients β_i leads to the following results (see Table 5.2). These results are based on the extended quality taxonomy resulting from the extensive subjective quality assessments only and, up to now, exclude characteristics based on the TR as extensively traced to parameters by Rittermann (see [26] and Chapter 4).

Table 5.2: Quality parameter per QF for the TR and respective QEs used.

QFs ***QEs***	incomplete, holey	blurred	contorted	stripes and lines	deformations, shape distortions	pixelated	flicker, jitter	color distortions, bad color quality	unlikely background	wrong proportions	slow, lagged	color spots, blemished	**s3DVQM**
A1-S1	4.18	0.23	2.67	0.00	0.25	0.60	0.07	0.06	0.04	0.25	0.44	0.77	**09.56**
A1-S2	5.31	0.24	2.42	0.00	0.34	0.34	0.25	0.03	0.04	0.45	0.46	0.44	**10.34**
A2-S1	4.50	0.22	2.99	0.00	0.19	0.40	0.14	0.04	0.05	0.28	0.52	0.60	**09.96**
A2-S2	4.25	0.26	3.62	0.00	0.49	0.33	0.06	0.01	0.03	0.33	0.38	0.74	**10.49**
A3-S1	5.80	0.24	2.39	0.00	0.53	0.48	0.11	0.03	0.09	0.35	0.38	0.36	**10.76**
A3-S2	5.83	0.27	3.05	0.00	0.49	0.53	0.15	0.02	0.03	0.18	0.39	0.08	**11.01**
A2-S3	4.23	0.21	0.00	0.09	3.02	0.28	0.95	0.19	0.04	0.36	0.09	0.36	**09.84**
A4-S4	0.00	0.24	2.43	0.69	3.70	0.51	0.25	0.14	0.00	0.67	0.11	0.45	**09.18**
A4-S5	0.44	0.20	2.38	1.17	0.72	0.43	0.84	0.12	0.00	0.84	0.13	0.42	**07.68**
A5-S4	0.60	0.16	0.47	0.70	2.46	0.59	0.65	0.16	0.05	1.92	0.05	0.58	**08.38**
A5-S5	1.46	0.19	1.42	1.22	0.60	0.63	0.97	0.05	0.06	0.94	0.06	0.24	**07.82**

These presented Subjective 3D Video Object Quality Metric (s3DVQM) values show that the QEs A3-S2, A3-S1, A2-S2 and A1-S2 have the highest QF influences and therefore may be the least successful TR combination in terms of the subjective perception of quality. Here, the QF incomplete, holey plays a crucial influencing role (because the individual QF value is higher (with 5.83, 5.80, 4.25 and 5.31) than for other QEs), followed by contorted (with values of 3.05, 2.39, 3.62 and 2.42). Comparing the values of the different QFs again leads to the result that incomplete, holey and contorted remain as the most severe factors overall.

Stripes and lines are less of an issue for the QEs (from A1-S1 to A3-S2) tested in Study 2. More important for these QEs is the QF slow and lagged. As already discussed in Chapter 4, this results from the test setup. Similar results are obtained for the QEs (from A2-S3 to A5-S5) tested in Study 3. In addition to the different used TR per QE, these s3DVQM values and their compositions also represent the influence of the entire usage process, including usage context, way of presentation, and way of judgement.

The advantage of these s3DVQM values is that they could be used to benefit the future TR of 3DVO. Taking these results into account, rules to benefit adaptive and scalable TRs and to guarantee a certain level of quality could be defined in a next step. These rules would then lead to the choice of filling algorithms or representation window size, etc., based on certain requirements such as available resources or severity of QE influences on certain QFs.Accounting for this model refinement of weighting by an expert results in the following refined extended quality taxonomies, based on Study 2 (see Figure 5.3) and Study 3 (see Figure 5.4). The expert judged by subjective assessment (see description of AHP in Saaty [212]), based on his knowledge and experience of creating the test stimuli (3DVO).

The differences described above between the results of Study 2 and Study 3 are shown in Figure 5.3 and Figure 5.4. Comparing the refined quality taxonomies with the Figures 5.1 and 5.2 demonstrates that the weighting has the necessary influence. Overall, the results are dominated by the subjective quality assessment. On the one hand, this assessment is influenced by the subjective quality assessment by test participants of a user-centered approach (see Study 3, in particular). On the other hand, it is weighted by an expert's subjective judgment, because 3DVOs still represent a rather new and uncommon visualization for end users. However, the subjective quality assessment is formalized into a figure s3DVQM that is intended to be comparable to the objective approach (3DVQM).

The presented extended taxonomies meet the predefined requirements to use information about the considered TR and accompanying QE as a basis and to take into account changes of the considered TR and, thus, changes of the QE. They also allow the presentation and the adequate usage of the knowledge about QF (derived from previous

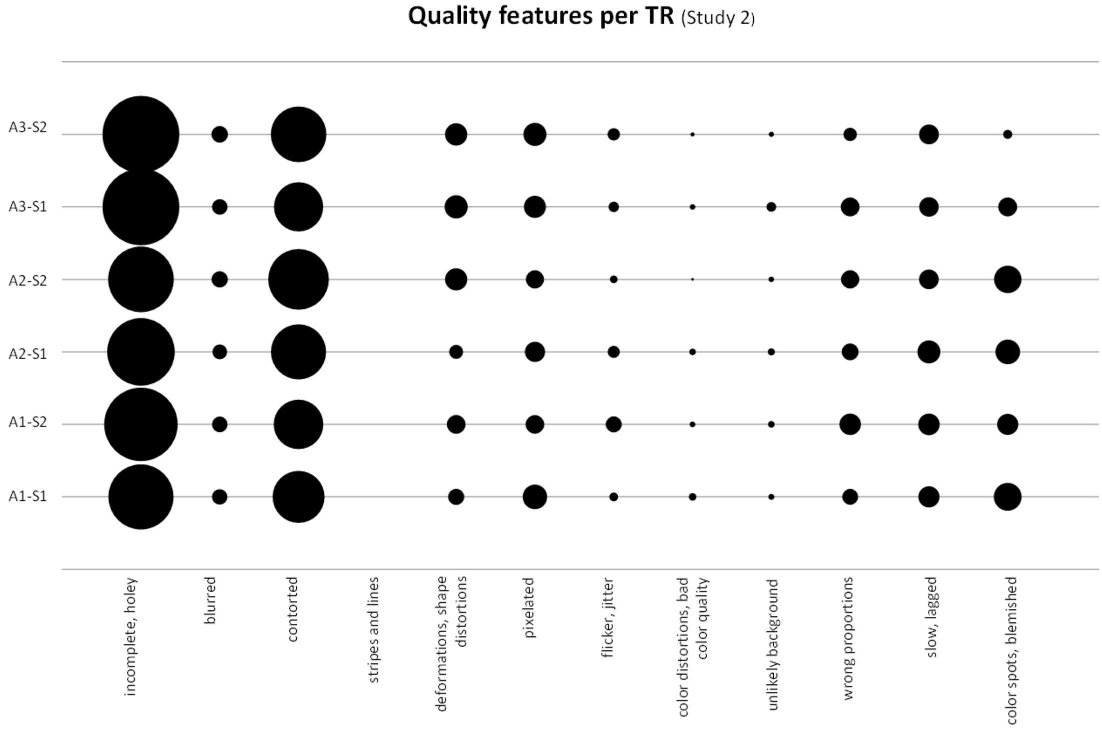


Figure 5.3: Refined quality taxonomy derived from Study 2.

subjective quality assessment, see Section 5.4).

This is shown by an exemplary measure, namely the s3DVQM adopted from Rittermann's 3DVQM [26]. This measure considers influences from changes of QE on QF and vice versa and should be complemented by a more detailed definition of QE influences after further 3DVO development. With this, the links between QE and QF could be displayed reliably. However, threshold definitions can be carried out and allow a more practicable integration framework related to adaptability and scalability. This can be supplemented with the usage of real-time capable and further developed systems, and with the knowledge about severe quality influencing QFs now available. For that reason the s3DVQM is further developed in terms of reprsenting (yet) unknown unknowns and unlinearities in the next subsection.

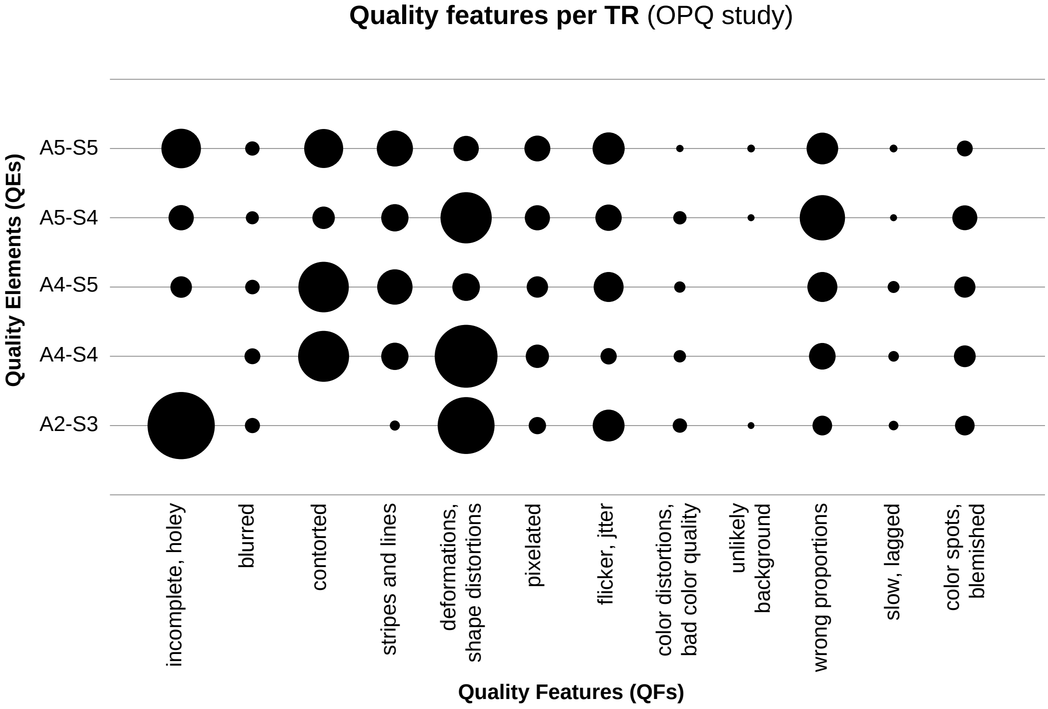


Figure 5.4: Refined quality taxonomy derived from Study 3.

5.3.4 Consideration of Nonlinearities and Unknowns

As the measure should be able to consider potential thresholds in the future, and the different kinds of influences on quality are in many cases nonlinear and may change over time as well, additional parameter, representing these issues, need to be included. One example for an unknown parameter might be *interacitvity* (or however it is named after defining the related QF and QE and these two perspectives) in the context of video communication. Nonlinearity is already considered implicitly by the bubbles' sizes in the quality taxonomy (see Figures 5.3 and 5.4). However, to work with more concrete information about the characteristics of QF:QE relation, these nonlinearities need to be considered explicitly. So it is required that the quality taxonomy allows consideration of recurrences in order to represent possible changes over time or changes from single (the relationship changing) influencing factors. These factors may be contextual or other define able influences on quality. Furthermore, e.g., based on other ways of TR as well as 3DVO usage, additional variables may occur. Hence, an approach towards formalizing the quality taxonomy is given in the next and final part of this thesis.

Coming back to the circle interlinking QoE with TR (as represented in Figure 1.2) and to the levels of produced and perceived quality (presented in Figure 1.5 in Section 1.1), a further development is proposed based on these (see Figures 5.5 and 5.6). This is simply based on available information provided in this work and further investigations towards potential relationships are required. Following flow chart in Figure 5.5 and slope field in Figure 5.6 intend the visualization of these potential relationships and may work as a starting point for further investigations thereon.

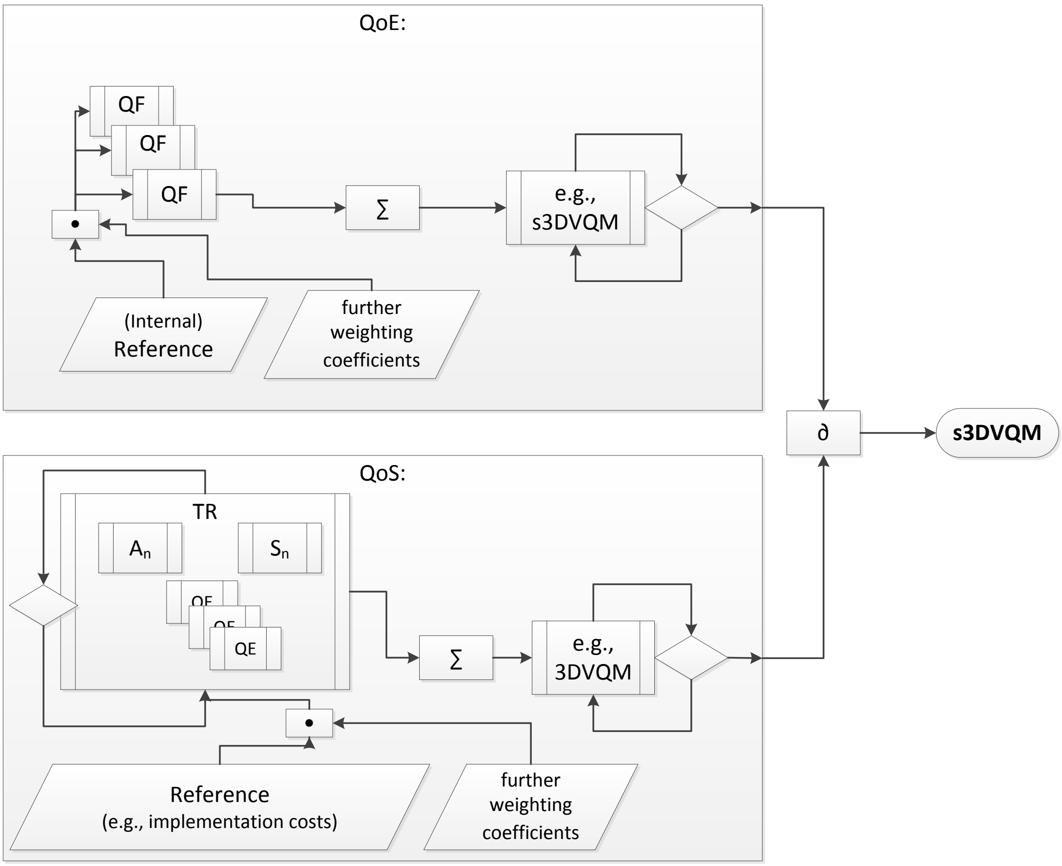


Figure 5.5: Flow Chart towards formalizing the quality taxonomy for a s3DVQM.

In Figure 5.5, the approach to separately define the single QE as well as the QF, including possible changes through iterations and references, is presented. It shows above the process to define QoE using weighted summations of every considered QF. These are weighted with known coefficients e.g., known influences by (internal) reference or usage context, which may change after several iterations and based on changes of the single influencing features. Furthermore, it shows below the process of defining QoS in a similar way. Namely, the process uses weighted summations of every considered QE as well and may change as well i.e., over time, or if an additional element has to be considered or another element is not important anymore. Bringing together the two parts i.e., above and below, e.g., by using a differential equation, leads to either a comparable measure (e.g., s3DVQM) or, by picking out single QF and QE, threshold values. This leads to the slope field in Figure 5.6.

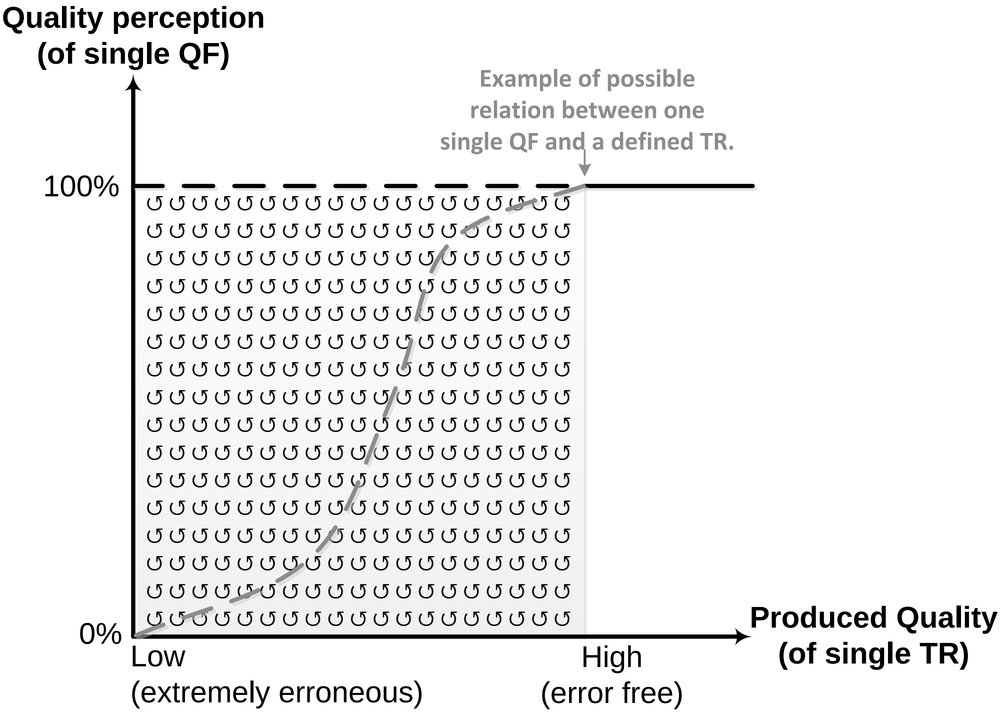


Figure 5.6: Slope Field towards formalizing the quality taxonomy for a s3DVQM.

Figure 5.6 shows the attempt to create a slope field in order to define useful functions or relationships based on information gathered (and represented as in Figure 5.5). Currently, instead of well defined slope directions, circles (↺) are used as place holder (The direction of the circle has no meaning here, as it could be the other way around, in any case based on the information used as basis. Therefore, circlearrows are used instead of an arrow towards any single direction). Once one flow chart for one single relation of QF to QE is defined a possible function (see example of sigmuid function in grey) could be differentiated.

Taking this as basis, a formula is still intended. However, the reasonableness of already formalizing a mathematical formula as a basis for a quality model is questionable with yet still unknown relations and influences. Hence, this first approach, for a subjective 3D video object quality metric which could become an adaptive subjective 3D video object quality metric s3DVQM, in a semi-mathematical and more descriptive way, is suggested for further investigations beyond the framework of this thesis. In other words, it is intended to differentiate the (weighted) quality influencing factors (QFs as well as QEs) and to map them into a multidimensional quality metric whichever direction the single influences have on the produced quality and perceived quality axis (see Figure 5.6). This could be used, for example, to define an activation function for further investigations using machine learning approaches. However, this has to be further validated in terms of reliable representation of the QoE and the right intensity of different influences and more importantly after a clear distinction of the then used basis that means, a construct including the TR its QEs and QFs leading to different priorities and therefore, to different magnitudes of influence.

5.4 Summary - Quality Taxonomy

This chapter presented and compared different available approaches to consider evaluation results within the TR processes used to create the investigated test stimuli. The conclusion is that, with the available results derived from the evaluation activities described in Chapter 4, the taxonomy approach introduced in [4] is most useful for the purpose of enhancing the TR processes of the investigated application, used within the Skalalgo3d project (see Section 2.3.2), at the current state of available TRs and by the current available knowledge of QFs.

The results presented in this chapter are based on the predefined use case 3DVO to support eye contact in video communication. Although the intention is to support eye contact, this topic was, finally, not mentioned by the test participants as a relevant quality influencing QF related to underlying QE in Study 3. Here, eye contact was not explicitly asked for beforehand when asking for the description of quality, as in Study 2, and was also not mentioned by the test participants themselves. In Study 2, eye contact was assessed separately too, but before the individual quality description of the test participants, and afterwards, was not mentioned by the test participants in their quality description. However, a possible correlation of the quality ratings and eye contact provided by the different QEs could be detected (see Sections 4.2.2 and 4.2.3).

The suggested quality taxonomy for scalable algorithms of 3D video objects functions as an overview of available relations between investigated QEs and the accompanying perceived QFs. In order to obtain a more viable description of the links between used TRs and perceived QFs, a AHP according to Saaty [212] was conducted after the creation of quality taxonomies from the available evaluation results and compared to a relevance ranking by an expert. In a further step, their weights and eigenvalues were included exemplarily into the 3DVQM introduced by Rittermann [26]. This is complemented by a suggestion on how to integrate the knowledge collected within this thesis to benefit the TR process of the investigated kinds of representations.

In sum, this work presents further knowledge about the influence of different TRs to create 3DVO and compares the presented results with the state-of-the-art of 3DVO creation and the future progress in this field. This knowledge is used to determine the resulting quality in a user-centered way and, thus, only for the presented usage context defined within the Skalalgo3d project. A PoC for investigated applications should be conducted using a real-time capable application and including further influences that may arise in a future step. To date, available links of the QFs to underlying QEs are mainly reduced and superficial, but user-centered and based on the subjective assessment of quality. Overall, the main results in this particular use case are the pre-defined quality influencing QFs.

The defined relevant QFs and this interlinking approach may be refined and complemented after having collected further knowledge concerning the usage of real-time capable applications, the integration of interaction and context-related factors, and regarding the possibility of threshold definition. This altogether may be validated with parametrization tools, for example, employing biofeedback as indicator.

6 Conclusion

6.1 Summary

The aim of this thesis is to provide a quality taxonomy based on subjective quality assessment for scalable algorithms of 3D video objects. The goal of the interlinking model, based on an extended taxonomy approach developed in this work, is to benefit the developing process of 3DVO and its assessed quality. Therefore, the inclusion of knowledge on subjective quality perception and features influencing quality in the perceptual domain is seen as advantageous. The definition of parameters representing quality that results in a metric is regarded as ideal.

It was noted that, although a 3DVQM exists (see [26]), a user-centered approach to the definition of QoE that includes more extensive subjective quality assessments is lacking. Thus, adequate and user-centered definitions of possible quality influencing QFs (e.g., blur, deformation, hole) have not been available, to date.

Hence, activities that can define QFs are presented in this work. This includes exploratory evaluations (i.e., within Study 1), in a first step, in order to explore possibly important QFs. In a second step, and considering changes in quality perception, due to further development and changes of related QEs, two laboratory evaluations (i.e., Study 2 and Study 3) were conducted. Based on these results, a list of QFs that can be considered for the overall quality definition was developed. It consists of 12 QFs. These are: incomplete/holey, blurred, contorted, stripes and lines, deformations/shape distortions, pixelated, flicker/jitter, color distortions/bad color quality, unlikely background, wrong proportions, slow/lagged, and color spots/blemished. These QFs, related to each evaluated QE, build the quality taxonomy that is based on the model introduced by Silzle [4], but include subjective quality assessment results of naïves and one expert, instead of only on expert evaluations. The distinction between different QFs, such as, for example, between contortions and deformations/shape distortions, has to be further investigated in the future. In this work, these quality describing attributes derived from a user centered approach lead to a different perception of these quality influencing features. However, this might not be the case in other use cases and for other related QEs. The same is probably also true for possible influences such as usage context, interaction and task, which have not been considered explicitly in the evaluations for the measure definition.

6.2 Contributions

Weighting factors were defined with the goal of benefitting the TR. Furthermore, an exemplary measure and a suggestion for the integration of evaluation results (or better: the knowledge about quality rating) into a scalable usage of algorithms was provided and represented by an exemplary measure definition. This was based on the pre-defined usage context within the Skalalgo3d project and involved TRs, leading to the definition of considered QEs.

The value of this proposal can be linked to the advantages given by (a) additional detected quality influencing features (i.e., QFs) defining the quality perception and their definition(s) provided by users, and (b) a systematization of the relations between the used QEs and perceived QFs given by the extended quality taxonomy. The term *extended* is necessary, because of the suggestion about how to weight and integrate the knowledge into the TR processes synthesizing an existing quality model.

This extended quality taxonomy is provided by the s3DVQM which makes use of the 3DVQM introduced by Rittermann [26], but its initial weighting coefficients are based on the subjective assessment of quality presented within this work and not on the theoretical lessons learned during the modeling of quality parameters.

6.3 Limitations

A paired-comparison of the various algorithms used (e.g., A1-S2, etc.) with the inclusion of possible end users has not been conducted to define thresholds. The reasons for this are multiple: A paired-comparison of the algorithms under test (eleven QEs in sum) would have required an overly large number of combinations if the focus had been on each of the 12 defined QFs. Consequently, approaches are available that reduce the duration of paired-comparison experiments in the context of visual quality assessment (e.g., Silverstein [213], or Mantel [214]).

However, these are approaches which assume that the quality perception can be represented within one dimension or in a linear fashion, which is not the case in the current investigation (see results in Chapters 4 and 5). At this point, one topic arises again: the differences and similarities between image fidelity and image quality, as introduced in Chapter 1 (see also [213]).

Another debatable point may be the weighting of the detected QFs by only one expert (the one who had performed the TR). This has to be further investigated and it has to be checked whether very different weights would result when asking other "‘experts"’, even though they may be less experienced. However, in order to propose the concept itself this approach to ask the only expert about the used TR seemed to be justifiable.

6.4 Future Work

In order to really integrate the collected knowledge into the TR processes in a very practical way, further developments related to the weighting structure, severity consideration, QE-QF links, and related adoption possibilities, as well as scalability are needed. Towards this goal, a more semantic approach that considers the presented results and/or machine learning approaches would be useful. In the future, more precision about physiological aspects of human vision could be used for parametrization issues. This could be, for example, a binocular eye tracking method, as presented within the BinoQ3D project that the author was involved in [22].

The author has suggested a way to overcome the lack of real-time capability within evaluation activities [23] that supports the useful approach of mental constructs (i.e., “imagine you are in a conversation with a friend/business partner by using a video communication application”), as used in Study 2. However, a further PoC that deals with relevant QFs in the context of 3DVO used in video communication systems has to be considered through the usage of a real-time capable application, in order to include influencing factors derived from the context of use, interaction, and from the form of transmission. Here, a possible predefined adoption of the quality by the user could be considered, e.g., by asking for the kind of device used or for the kind of access.

Additionally, further fine-tuning concerning the weighting of the severity of the QFs could be useful. Therefore, a large scale evaluation similar to AHP or a threshold evaluation of test stimuli showing different severities of different QFs per QE would be helpful. However, to date, it is not always possible to control the impact of a certain TR process on the resulting quality. Thus, it was not yet possible to provide test stimuli with controlable nuances of quality influencing QFs. When this is possible, a further QF refinement should be undertaken. Furthermore, the concept itself should be examined under another use case, in order to refine it and to better indicate its advantages and disadvantages. In doing so, more attention could be paid to the topic of *interaction* as an influencing factor, as is being considered within newly developed approaches towards the interaction with 3DVO on TV [21]. The author proposes a combination of a QF defining investigation followed by a pairwise comparison with end users for different contexts, in order to fill the required values for a quality defining measure, including weighting factors.

6.5 Final Remarks

This work shows that the approach of subjective assessment of scalable algorithms of 3D video objects depends on the available stimuli and application status. Nevertheless, quality differentiation is given within a user-centered approach, even in the low quality range and within early development phases. The effort for achieving this will undoubtedly remain. The need for more information about thresholds and intensities of QF perception is also unquestionable (i.e., usage of method of limits, or continuous scaling experiments, as described in the ITU-T BT.500 [36]).

The proposal for an iterative, interlinking model in the form of weighting factors and defined relations is seen as contribution to the definition of the quality perception of related space(s) within the ARCU approach [14], or support the definition of an activation function for further investigations using machine learning approaches. It can also be seen as contribution to the roadmap introduced by Rogowitz and Goodman [13] to include a *human-in-the-loop*.

Bibliography

[1] C. Weigel and N. Treutner, “Establishing eye contact for home video communication using stereo analysis and free viewpoint synthesis,” in *Proceedings of SPIE*, vol. 8290, 2012, pp. 829 003–829 003–9, online available at http://dx.doi.org/10.1117/12.906594, last seen on 22nd Nov. 2013.

[2] C. Weigel, “Ein Beitrag zur Entwicklung von Methoden zur Stereoanalyse und Bildsynthese im Anwendungskontext der Videokommunikation,” Ph.D. dissertation, Technische Universität Ilmenau, Germany, 2014.

[3] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley, 1995.

[4] A. Silzle, *Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey*, ser. Berichte aus der Kommunikationstechnik. Shaker, 2007.

[5] U. Jekosch, *Voice and Speech Quality Perception. Assessment and Evaluation.* Springer, Berlin, Germany, 2005.

[6] U. Kühhirt and M. Rittermann, *Interaktive audiovisuelle Medien.* Carl Hanser Verlag, München, GER, 2007.

[7] J. Foss, B. Malheiro, and J. C. Burguillo, “Personalization of Networked Video,” in *Adjunct Proceedings of EuroITV 2012*, 2012, p. 132.

[8] R. Mekuria, M. Sanna, S. Asioli, E. Izquierdo, D. C. A. Bulterman, and P. Cesar, “A 3D tele-immersion system based on live captured mesh geometry,” in *Proceedings of the 4th ACM Multimedia Systems Conference*, ser. MMSys ’13. New York, NY, USA: ACM, 2013, pp. 24–35, online available at http://doi.acm.org/10.1145/2483977.2483980, last seen on 22nd Nov. 2013.

[9] M. Pollefeys, K. Sing Bing, S. Gregory G., C. Kurt, and D. Paul, "Panel Session on Visual Scene Representation," in *3D Structure from Images - SMILE 2000, Springer Lecture notes in computer science*, ser. Lecture Notes in Computer Science, M. Pollefeys, L. v. Gool, A. Zisserman, and A. Fitzgibbon, Eds. Springer, Berlin, GER, 2001, vol. 2018, pp. 161–169, second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments Dublin, Ireland, July 12, 2000, Revised Papers.

[10] C. Weigel, S. Schwarz, T. Korn, and M. Wallebohr, "Interactive free viewpoint video from multiple stereo," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, may 2009, pp. 1–4.

[11] S. Winkler, *Digital Video Quality: Vision Models and Metrics.* Wiley, 2005.

[12] S. Jumisko-Pyykkö, D. Strohmeier, T. Utriainen, and K. Kunze, "Descriptive quality of experience for mobile 3D video," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ser. NordiCHI '10. New York, NY, USA: ACM, 2010, pp. 266–275, online available at http://doi.acm.org/10.1145/1868914.1868947.

[13] B. E. Rogowitz and A. Goodman, "Integrating human- and computer-based approaches to feature extraction and analysis," in *Proceedings of SPIE*, vol. 8291, 2012, pp. 82 910W–82 910W–11, online available at http://dx.doi.org/10.1117/12.915887, last seen on 22nd Nov. 2013.

[14] L. Skorin-Kapov and M. Varela, "A Multi-Dimensional View of QoE: the ARCU Model," in *Proceedings of 35th MIPRO*, ser. MIPRO, May 2012.

[15] D. Tobian, "Konzeptionierung einer Anwendungssimulation zur Pre-Prototypischen Evaluierung," Media project, Technische Universität Ilmenau, Germany, 2012.

[16] J. G. Beerends and F. E. De Caluwe, "The influence of video quality on perceived audio quality and vice versa," *Journal of the Audio Engineering Society*, vol. 47(5), pp. 355–362, 1999.

[17] S. Kepplinger, "Subjective Quality Assessment of Free Viewpoint Video Objects," Adjunct Proceedings of EuroITV 2011, online available at www.scribd.com/doc/

58835687/EUROITV2011-Adjunct-Proceedings, last seen on 22nd Nov. 2013, pp. 47–50, 2011.

[18] S. Kepplinger and C. Weigel, "Towards a Model to Interlink Quality of Experience with Algorithm Development," in *Proceedings of International Conference on Consumer Electronics in Berlin (ICCE-Berlin)*, 2011, pp. 239–242.

[19] S. Kepplinger, "Linking quality assessment of free-viewpoint video objects up with algorithm development," in *Proceedings of SPIE*, vol. 8293, 2012, pp. 82 930U–82 930U–12, online available at http://dx.doi.org/10.1117/12.906582, last seen on 22nd Nov. 2013.

[20] ——, "Profiling User Perception of Free Viewpoint Video Objects in Video Communication," Adjunct Proceedings of EuroITV 2012, online available at www.euroitv2012.org/AdjProc_EuroITV2012.pdf, last seen on 22nd Nov. 2013, pp. 63–66, 2012.

[21] S. Kepplinger, J. Foss, I. Cristea, Alexandra, and J. Liebetrau, "Roadmap for a comprehensive Evaluation Approach on QoE of interactive and personalized TV," Adjunct Proceedings of EuroITV 2013, online available at http://www.euro-itv.org/program/proceedings, last seen on 22nd Nov. 2013, pp. 166–169, 2013.

[22] S. Kepplinger, F. Hofmeyer, and M. Gründl, "Development of a Binocular Eye Tracking System for Quality Assessment of S3D Representations," in *Quality of Multimedia Experience (QOMEX), 2013 Fifth International Workshop on*, 2013, pp. 124–125.

[23] S. Kepplinger and D. Tobian, "Wizard of Oz Approach for a 3DVO Video Communication System," in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013), 2nd–4th September 2013, Vienna, Austria*, 2013, pp. 157–160.

[24] S. Kepplinger and N. Hottong, "Quality evaluation of stereo 3DTV systems with open profiling of quality," *Proceedings of SPIE*, vol. 9014, no. 9014-46, pp. 901 419–901 419–11, 2014, (Online available at http://dx.doi.org/10.1117/12.2038981, last seen on 29th Feb. 2016).

[25] S. Kepplinger, “Linking Quality Features and Quality Elements by defining an extended 3DVO Quality Taxonomy Model,” in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on.* IEEE, Sept 2014, pp. 41–46.

[26] M. Rittermann, “Zur Qualitätsbeurteilung von 3DVideoobjekten,” Ph.D. dissertation, Technische Universität Ilmenau, Germany, 2007.

[27] M. Pollefeys, L. v. Gool, A. Zisserman, and A. Fitzgibbon, Eds., *3D Structure from Images - SMILE 2000*, ser. Series: Lecture Notes in Computer Science. Springer, 2001, vol. 2018, second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments Dublin, Ireland, July 12, 2000, Revised Papers.

[28] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[29] A. Smolic, K. Müller, P. Merkle, M. Kautzner, and T. Wiegand, “3D video objects for interactive applications,” in *13th European Signal Processing Conference (EUSIPCO 2005)*, ser. EUSIPCO. Lausanne, Switzerland: EURASIP, 2005, pp. 640–644.

[30] ITU, *Definitions of terms related to quality of service (ITU-T E.800)*, International Telecommunication Union, Geneva, 2008.

[31] Nokia, *Quality of Experience (QoE) of Mobile Services: Can it be Measured and Improved?*, 2004, white Paper, online available at http://www.nokia.com/NOKIA_COM_1/Operators/Downloads/Nokia_Services/whitepaper_qoe_net.pdf,Institution:Nokia, last seen on 23rd June 2010.

[32] *Qualinet White Paper on Definitions of Quality of Experience*, 1st ed., online available, European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), June 2012, output from the Dagstuhl seminar 12181. Online available at http://www.qualinet.eu/images/stories/whitepaper_v1.1_dagstuhl_output.pdf, last seen on 22nd Nov. 2013.

[33] ITU, *Vocabulary and Effects of Transmission Parameters on Customer Opinion of Transmission Quality (ITU-T P.10/G.100)*, International Telecommunication Union, Geneva, 2008.

[34] ISO, *Ergonomics of human system interaction (9241), Part 210: Human-centred design for interactive systems*, International Organization for Standardization (ISO), 2010.

[35] *User Experience White Paper. Bringing clarity to the concept of user experience*, Allaboutux.org, 2011, result from Dagstuhl Seminar on Demarcating User Experience 2010, online available at http://www.allaboutux.org/files/UX-WhitePaper.pdf, last seen on 22nd Nov. 2013.

[36] ITU, *Methodology for the subjective assessment of the quality of television pictures (ITU-T BT.500-13)*, International Telecommunication Union, Geneva, 2012.

[37] D. Geerts, K. De Moor, I. Ketykó, A. Jacobs, J. Van den Bergh, W. Joseph, L. Martens, and L. De Marez, "Linking an Integrated Framework with Appropriate Methods for Measuring QoE," in *Second International Workshop on Quality of Multimedia Experience (QoMEX 2010).* IEEE, 2010, pp. 158–163.

[38] O. Schreer, P. Kauff, and T. Sikora, Eds., *3D Video communication: Algorithms, concepts and real-time systems in human centered communication.* John Wiley and Sons, Ltd., UK, 2005.

[39] S. Jumisko-Pyykkö, V. K. Malamal Vadakital, and M. M. Hannuksela, "Acceptance Threshold: A Bidimensional Research Method for User-Oriented Quality Evaluation Studies," *International Journal of Digital Multimedia Broadcasting*, vol. 2008, no. 712380, p. 20, 2008, online available at http://dx.doi.org/10.1155/2008/712380, last seen on 22nd Nov. 2013.

[40] R. McDonnell and M. Breidt, "Face reality: Investigating the uncanny valley for virtual faces," in *ACM SIGGRAPH ASIA 2010 Sketches*, ser. SA '10. New York, NY, USA: ACM, 2010, pp. 41:1–41:2. [Online]. Available: http://doi.acm.org/10.1145/1899950.1899991

[41] J. Korhonen, U. Reiter, and Y. Junyong, "Subjective comparison of temporal and quality scalability," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, 2011, pp. 161–166.

[42] S. Foy, "Quality of Experience assurance for next generation video networks," Presentation at 25th Fernseh- und Kinotechnische Gesellschaft (FKTG) meeting, talk number 29, 2012.

[43] F. I. for Digital Media Technology IDMT, “A/V Analyzing Toolbox,” projects and products website, online available at http://www.idmt.fraunhofer.de/en/institute/projects_products/a_d/av_analyzing_toolbox.html, last seen on 19th September 2016, 2016.

[44] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, “3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards,” in *Multimedia and Expo, 2006 IEEE International Conference on*, july 2006, pp. 2161–2164.

[45] ISO, *Information Technology - Generic Coding of Audio-Visual Objects (MPEG-4)(ISO 14496)*, 1998.

[46] T. U. I. Institute of Media Technology, “IAVAS - Interactive AudioVisual Application Systems: 3D Video Object,” Technische Universität Ilmenau, Germany, Tech. Rep., 2004, also available at http://imt.rz.tu-ilmenau.de/iavas/i3d/, last seen on 22nd Nov. 2013.

[47] A. Smolic, H. Kimata, and A. Vetro, “Development of mpeg standards for 3d and free viewpoint video,” pp. 60 160R–60 160R–12, 2005. [Online]. Available: http://dx.doi.org/10.1117/12.631192

[48] A. Smolic and H. Kimata, *Applications and Requirements for 3DAV. ISO/IEC JTC 1/SC 29/WG 11 (Doc. N5877)*, July 2003.

[49] I. Feldmann and et al., “Immersive Multi-User 3D Video Communication,” in *Proceedings of International Broadcast Conference (IBC 2009)*, 2009.

[50] ISO, *Information technology - Coding of audio-visual objects - Part 16: Animation Framework eXtension (AFX). ISO/IEC 14496-16:2011*, November 2011.

[51] ——, *Use Cases and Requirements on Free-viewpoint Television (FTV), MPEG2013/N14104*, 2013.

[52] ISO/IEC, *Information technology - MPEG video technologies - Part 3: Representation of auxiliary video and supplemental information. ISO/IEC 23002-3:2007*, 2007.

[53] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Efficient prediction structures for multiview video coding,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1461–1473, Nov 2007.

[54] S. B. Kang, “Survey of image-based rendering techniques,” in *Proceedings of SPIE*, vol. 3641, 1998, pp. 2–16, online available at http://dx.doi.org/10.1117/12.333774, last seen on 22nd Nov. 2013.

[55] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing.* MIT Press, 1991, pp. 3–20.

[56] J.-R. Ohm and K. Müller, “Incomplete 3-D multiview representation of video objects,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 2, pp. 389–400, mar 1999.

[57] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH ’96. New York, NY, USA: ACM, 1996, pp. 31–42, online available at http://doi.acm.org/10.1145/237170.237199, last seen on 22nd Nov. 2013.

[58] F. Fan, “Analyse von Qualitätsparametern von 3D-Videoobjekten,” Diploma thesis, Technische Universität Ilmenau, Germany, 2008.

[59] K. Müller, *Rekonstruktion und Codierung von Free Viewpoint Video.* Mensch und Buch Verlag Berlin, GER, 2006.

[60] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” *Int. J. Comput. Vision*, vol. 47, no. 1–3, pp. 7–42, Apr. 2002, online available at http://dx.doi.org/10.1023/A:1014573219977, last seen on 22nd Nov. 2013.

[61] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, “Free-viewpoint video of human actors,” in *ACM SIGGRAPH 2003 Papers*, ser. SIGGRAPH ’03. New York, NY, USA: ACM, 2003, pp. 569–577, online available at http://doi.acm.org/10.1145/1201775.882309, last seen on 22nd Nov. 2013.

[62] A. Telea, “An Image Inpainting Technique Based on the Fast Marching Method,” *Graphics Tools*, vol. 9, pp. 23–34, 2004.

[63] B. Cyganek and J. P. Siebert, *An Introduction to 3D Computer Vision Techniques and Algorightms.* John Wiley and Sons, Ltd., UK, 2009.

[64] A. Geiger, M. Roser, and R. Urtasun, “Efficient Large-Scale Stereo Matching,” in *Computer Vision–ACCV 2010*, ser. Lecture Notes in Computer Science, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Springer Berlin Heidelberg, 2011, vol. 6492, pp. 25–38.

[65] A. A. Alatan, Y. Yemez, U. Güdükbay, X. Zabulis, K. Müller, C. E. Erdem, C. Weigel, and A. Smolic, “Scene Representation Technologies for 3DTV: A Survey,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1587–1605, nov. 2007.

[66] O. Grau, G. Thomas, A. Hilton, J. Kilner, and J. Starck, “iview: A Robust Free-Viewpoint Video System for Sport Scenes,” BBC, White Paper WHP 149, Tech. Rep., 2007, also available at http://www.bbc.co.uk/rd/projects/iview, last seen on 22nd Nov. 2013.

[67] R. Yang and Z. Zhang, “3D Video,” Microsoft Research, Tech. Rep. MSR-TR-2001-119, December 2001, also available at http://research.microsoft.com/pubs/69938/tr-2001-119.pdf, last seen on 22nd Nov. 2013.

[68] LiberoVision, “Sports Broadcasting. Redefined.” Online available at http://www.vizrt.com/products/viz_libero/, last seen on 22nd Nov. 2013, 2007–2011.

[69] tvmitschnitte, “Das neue 3D-Spielanalyse-Tool vom ZDF,” Online available at http://www.youtube.com/watch?v=CRZMzvdZDHc, last seen on 22nd Nov. 2013, 2008.

[70] L. Ballan, “Unstructured Video-Based Rendering: Interactive Exploration of Casually Captured Videos,” Online available at https://youtu.be/gqSLsPrA7RE, last seen on 2nd Jan. 2017, 2010.

[71] M. Argyle and M. Cook, *Gaze and mutual gaze.* Cambridge University Press, New York, USA, 1976.

[72] N. Arif, H. Karim, H. Mohamad, A. Sali, and M. Mokhtar, “User perception evaluations of 3D video telephony as compared to actual experience for face-to-face immersive communications,” in *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*, june 2011, pp. 602–606.

[73] T. Korn, "Kalibrierung und Blickrichtungsanalyse für ein 3D Videokonferenzsystem," Diploma thesis, Technische Universität Ilmenau, Germany, 2009.

[74] C. Kuster, T. Popa, J.-C. Bazin, C. Gotsman, and M. Gross, "Gaze Correction for Home Video Conferencing," *ACM Trans. Graph. (Proceedings of ACM SIGGRAPH ASIA)*, vol. 31, no. 6, pp. 174:1–174:6, 2012.

[75] R. Stokes, Rembert, "Human Factors and Appearance Design Considerations of the Modern PICTUREPHONE Station Set," *IEEE Transactions on Communication Technology*, vol. Com-17, no. 2, 1969.

[76] W.-C. Chen, H. Towels, L. Nyland, G. Welch, and H. Fuchs, "Toward a compelling sensation of telepresence: Demonstrating a portal to a distant (static) office," in *Visualization 2000*, 2000, pp. 327–333.

[77] P. Eisert, "Immersive 3-D Video Conferencing: Challenges, Concepts, and Implementations," 2003, sPIE Visual Communications and Image Processing (VCIP), Lugano, Switzerland.

[78] F. Solina and R. Ravnik, "Fixing Missing Eye-Contact in Video Conferencing Systems," in *Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces, June 27–30, 2011, Cavtat, Croatia*, 2011.

[79] D. Murayama, K. Kimura, T. Hosaka, T. Hamamoto, N. Shibuhisa, S. Tanaka, S. Sato, and S. Saito, "Virtual view image synthesis for eye-contact in TV conversation system," in *Proceedings of SPIE*, vol. 7526, 2010, pp. 75 260B–75 260B–10, online available at http://dx.doi.org/10.1117/12.839121, last seen on 22nd Nov. 2013.

[80] R. van Eijk, A. Kuijsters, K. Dijkstra, and W. A. IJsselsteijn, "Human sensitivity to eye contact in 2D and 3D videoconferencing," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, june 2010, pp. 76–81.

[81] M. F. Cohen, R. A. Colburn, and S. M. Drucker, "The Role of Eye Gaze in Avatar Mediated Conversational Interfaces," Microsoft Research, Tech. Rep. MSR-TR-2000-81, July 2000, avatar Mediated Conversational Interfaces.

[82] R. V. Kollarits, C. Woodworth, J. Ribera, and R. Gitlin, *An eye contact camera/display system for videophone applications using a conventional direct-view LCD.* SID 1995, Digest, 1995.

[83] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. V. Moere, and O. Staadt, "blue-c: a spatially immersive display and 3D video portal for telepresence," in *ACM SIGGRAPH 2003 Papers*, ser. SIGGRAPH '03. New York, NY, USA: ACM, 2003, pp. 819–827, online available at http://doi.acm.org/10.1145/1201775.882350, last seen on 22nd Nov. 2013.

[84] I. Feldmann and O. Schreer, "Virtual Eye Contact Engine," Online available at http://www.hhi.fraunhofer.de/fields-of-competence/image-processing/solutions/capture/virtual-eye-contact-engine.html, last seen on 22nd Nov. 2013, 2011.

[85] M. Hellfritzsch, "Untersuchungen und Vergleich zur Definition der Nutzungsumgebung und des Einsatzes von Video Kommunikation im professionellen und im privaten Bereich," Master thesis, Technische Universität Ilmenau, Germany, 2012.

[86] K. Brandenburg (proposer), "Skalierbare Algorithmen für 3D Videoobjekte unter Berücksichtigung subjektiver Qualitätsfaktoren (Skalalgo3d), Grant BR 1333/8-1," 2009–2012, this work was supported by a grant of the German research foundation (DFG).

[87] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," release and documentation website, online available at http://www.vision.caltech.edu/bouguetj/calib_doc/, last seen on 18th May 2016, 2010.

[88] ISO, *Quality management systems - Fundamentals and vocabulary (9000)*, 2005.

[89] J. Ludewig and H. Lichter, *Software Engineering - Grundlagen, Menschen, Prozesse, Techniken*, 2nd ed. dpunkt.verlag, Germany, 2010.

[90] N. Zacharov, "Statistical experimental design - statistical analysis," Lecture notes at 1st Qualinet Summer School on Quality Assessment, 2012.

[91] D. S. Hands, "A basic multimedia quality model," *Multimedia, IEEE Transactions on*, vol. 6, no. 6, pp. 806–816, 2004.

[92] H. de Ridder and S. Endrikhovski, "Image Quality is FUN: Reflections on Fidelity, Usefulness and Naturalness," *SID Symposium Digest of Technical Papers*, vol. 33, no. 1, pp. 986–989, 2002, 33.1: Invited Paper. Online available at http://dx.doi.org/10.1889/1.1830949, last seen on 22nd Nov. 2013.

[93] P. G. Engeldrum, “A theory of image quality: The image quality circle,” *Journal of Imaging Science and Technology*, vol. 48, no. 5, pp. 447–457, 2004.

[94] F. Fechter, *Zur Beurteilung der Qualität komprimierter Bildfolgen.* VDI Verlag GmbH Düsseldorf, GER, 2000, habilitationsschrift an der Technischen Universität Braunschweig.

[95] S. Jumisko-Pyykkö, “User Cented Quality of Experience,” Ph.D. dissertation, Tampere University of Technology, Tampere, Finland, 2011.

[96] K. u. R. Laghari, N. Crespi, and K. Connelly, “Toward total quality of experience: A QoE model in a communication ecosystem,” *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 58–65, 2012.

[97] K. Diepold, “The Quest for a Definition of Quality of Experience,” Qualinet Newslet, October 2012, online available at http://www.qualinet.eu/images/stories/Qualinet-Newslet-Vol-2-2012.pdf, last seen on 19th Aug. 2014, pp. 2–8, 2012.

[98] J. Roufs and H. Bouma, “Towards linking perception research and image quality,” in *Proceedings of SID*, vol. 21/3, 1980, pp. 247–270.

[99] H. Coolican, *Research Methods and Statistics in Psychology*, 5th ed. Hodder Education, London, UK, 2009.

[100] DIN, *Concepts for quality management - Part 11: Supplement to DIN EN ISO 9000:2005*, DIN, the German Institute for Standardization, May 2008.

[101] E. B. Goldstein, *Sensation and Perception.* States of America: Wadsworth, USA, 2002.

[102] U. Neisser, *Cognition and Reality, Principles and Implications of Cognitive Psychology.* W.H. Freeman and Company, San Francisco, USA, 1976.

[103] S. Möller, B. Belmudez, M.-N. Garcia, C. Kühnel, A. Raake, and B. Weiss, “Audiovisual quality integration: Comparison of human-human and human-machine interaction scenarios of different interactivity,” in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, june 2010, pp. 58–63.

[104] D. Strohmeier, "Open profiling of quality: a mixed methods research approach for audiovisual quality evaluations," Ph.D. dissertation, Technische Universität Ilmenau, Germany, 2011, online available at http://www.db-thueringen.de/servlets/DocumentServlet?id=19609, last seen on 11th Sept. 2014.

[105] J. C. Baird, *Sensation and judgment: Complementary theory of psychophysics.* Lawrence Erlbaum Ass., Mahwah, New Jersey, USA, 1997.

[106] M. P. Hollier and R. M. Voelcker, "Towards a multi-modal perceptual model," *BT Technology Journal*, vol. 15, no. 4, pp. 163–172, Oct. 1997, online available at http://dx.doi.org/10.1023/A:1018695832358, last seen on 22nd Nov. 2013.

[107] M. P. Hollier, A. N. Rimell, D. S. Hands, and R. M. Voelcker, "Multi-modal Perception," *BT Technology Journal*, vol. 17, no. 1, pp. 35–46, Jan. 1999, online available at http://dx.doi.org/10.1023/A:1009666623193, last seen on 22nd Nov. 2013.

[108] P. G. Engeldrum, *Psychometric Scaling: A toolkit for imaging systems development.* Imcotec Press, 2000.

[109] H. de Ridder, "Cognitive issues in image quality measurement," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 47–55, 2001, online available at http://dx.doi.org/10.1117/1.1335529, last seen on 22nd Nov. 2013.

[110] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review of Neuroscience*, vol. 2004, pp. 649–677, 2004.

[111] S. Y. Chen, G. Ghinea, and R. D. Macredie, "A cognitive approach to user perception of multimedia quality: An empirical investigation," *Int. J. Hum.-Comput. Stud.*, vol. 64, no. 12, pp. 1200–1213, Dec. 2006, online available at http://dx.doi.org/10.1016/j.ijhcs.2006.08.010, last seen on 22nd Nov. 2013.

[112] A. Kayser, "Modellierung der Bildqualität datenkomprimierter stereoskopischer Fernsehbilder," Ph.D. dissertation, Technische Universität München, 1995.

[113] E. Balcetis and G. Lassiter, Daniel, Eds., *Social Psychology of Visual Perception.* Psychology Press, East Sussex, UK, 2010.

[114] K. Kunze, J. Liebetrau, and T. Korn, “Evaluating Spatial Congruency of 3D Audio and Video Objects,” in *Audio Engineering Society Convention 132*, Apr 2012, online available at http://www.aes.org/e-lib/browse.cfm?elib=16317, last seen on 22nd Nov. 2013.

[115] E. Britannica, “human sensory reception,” Online available at http://www.britannica.com/EBchecked/topic/534831/human-sensory-reception, last seen on 22nd Nov. 2013, 2013.

[116] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design - Beyond Human-Computer Interaction*, 3rd ed. Wiley, 2012.

[117] G. Simmel, “The Sociology of Sociability,” *American Journal of Sociology*, vol. 55, no. 3, pp. 254–261, 1949, translated by and Hughes, Everett C., online available at http://www.jstor.org/stable/2771136, last seen on 22nd Nov. 2013.

[118] V. Aaltonen, J. Takatalo, J. Häkkinen, M. Lehtonen, G. Nyman, and M. Schrader, “Measuring Mediated Communication Experience,” in *Proceedings of the First International Workshop on Quality of Multimedia Experience (QOMEX), July 29–31, 2009, San Diego, CA, USA*, ser. 1, no. 104, 2009.

[119] M. Fiedler, S. Möller, and P. Reichl, “Quality of Experience: From User Perception to Instrumental Metrics (Dagstuhl Seminar 12181),” *Dagstuhl Reports*, vol. 2, no. 5, pp. 1–25, 2012, online available at http://drops.dagstuhl.de/opus/volltexte/2012/3598, last seen on 22nd Nov. 2013.

[120] Y. Pitrey and H. Hlavacs, “You look blocky, is everything alright? Influence of video distortions on facial expression recognition & quality,” in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems*, ser. PQS, September 2013.

[121] Dudenredaktion, Ed., *Das Fremdwörterbuch, 10. Auflage.* Bibliographisches Institut GmbH, Mannheim, GER, 2010.

[122] D. D. Hoffman, *Visual Intelligence: How We Create What We See.* WW Norton, New York, USA, 1998.

[123] A. Watson and M. A. Sasse, “Evaluating Audio and Video Quality in Low-Cost Multimedia Conferencing Systems,” *Interacting with Computers*, vol. 8, pp. 255–275, 1996.

[124] S. Winkler and C. Faller, "Audiovisual quality evaluation of low-bitrate video," in *Proceedings of SPIE*, vol. 5666, 2005, pp. 139–148.

[125] ——, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 973–980, 2006.

[126] H. Knoche, J. McCarthy, and M. A. Sasse, "How Low Can You Go? The Effect of Low Resolutions on Shot Types in Mobile TV," *Multimedia Tools and Applications*, vol. 36, pp. 145–166, 2008.

[127] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bitrate videos," *IEEE Transactions on Multimedia*, vol. 10, no. 7, pp. 1316–1324, 2008.

[128] A. Ninassi, O. L. Meur, P. L. Callet, D. Barba, and A. Tirel, "Task impact on the visual attention in subjective image quality assessment," in *Proceedings of EUSIPCO-06*, 2006.

[129] A. Parducci and D. H. Wedell, "The category effect with rating scales: Number of categories, number of stimuli, and method of presentation," *Journal of Experimental Psychology: Human perception and performance*, vol. 12, no. 4, pp. 496–516, 1986.

[130] T. Naes and E. Risvik, Eds., *Multivariate Analysis of Data in Sensory Science.* Elsevier Science B.V., Amsterdam, The Netherlands, 1996, vol. 16.

[131] F. Speranza, F. Poulin, R. Renaud, M. Caron, and J. Dupras, "Objective and subjective quality assessment with expert and non-expert viewers," in *Proceedings of the Second International Workshop of Quality of Multimedia Experience: QoMEX 2010*, 2010, pp. 46–51.

[132] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. John Wiley, UK, 1982.

[133] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," in *Digital images and human vision*, A. B. Watson, Ed. Cambridge, MA, USA: MIT Press, 1993, pp. 179–206, online available at http://dl.acm.org/citation.cfm?id=197765.197783, last seen on 22nd Nov. 2013.

[134] J. Lubin and D. Fibush, “SarnoffJND vision model,” ANSI T1 Standards Committee, Tech. Rep., 1997, t1A1.5 Working Group, Document Nr. 97–612.

[135] C. J. van den Branden Lambrecht, “Perceptual Models and Architectures for Video Coding Applications,” Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 1996.

[136] P. Lindh and C. J. Van Den Branden Lambrecht, “Efficient spatio-temporal decomposition for perceptual processing of video sequences,” in *Image Processing, 1996. Proceedings., International Conference on*, vol. 3, sep 1996, pp. 331–334.

[137] M. A. Masry and S. S. Hemami, “A metric for continuous quality evaluation of compressed video with severe distortions, Signal Processing,” *Image Comm*, vol. 19, pp. 133–146, 2004.

[138] A. B. Watson, Q. J. Hu, J. F. McGowan III, and J. B. Mulligan, “Design and performance of a digital video quality metric,” in *Proceedings of SPIE Human Vision and Electronic Imaging, San Jose, CA*, vol. 3644, 1999, pp. 168–174.

[139] S. Wolf and M. Pinson, *Video Quality Measurement Techniques.* NTIA Report, Boulder, USA, 2002.

[140] VQEG, “Final report on the validation of objective models of video quality assessment,” Video Quality Experts Group (VQEG), Tech. Rep., 2000, also available at http://www.its.bldrdoc.gov/vqeg/reports.aspx, last seen on 22nd Nov. 2013.

[141] ITU, *Subjective Video Quality Assessment Methods for Multimedia Applications (ITU-T P.910)*, International Telecommunication Union, Geneva, 2008.

[142] M. Solh and G. AlRegib, “MIQM: A Novel Multi-View Images Quality Measure,” in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, 2009, pp. 186–191.

[143] P. Gastaldo, R. Zunino, H. Ingrid, and V. Elena, “Objective Quality Assessment of Displayed Images by using Neural Networks,” *Signal Processing: Image Communication*, vol. 20, no. 643, pp. 643–661, 2005, online available at http://atc.umh.es/gatcom/bin/oqam/Referencias/Gastaldo_Zunino_Heynderickx_Vicario_2005.pdf, last seen on 22nd Nov. 2013.

[144] S. Winkler, “On the Properties of Subjective Ratings in Video Quality Experiments,” in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, 2009, pp. 139–144.

[145] M. Pinson and S. Wolf, “Comparing subjective video quality testing methodologies,” in *SPIE Video Communications and Image Processing Conference*, 2003, pp. 8–11.

[146] F. Kozamernik, P. Sunna, E. Wyckens, and D. I. Pettersen, “Subjective quality of internet video codecs - phase 2 evaluations using SAMVIQ,” European Broadcasting Union, EBU Technical Review, Tech. Rep., 2005.

[147] F. Pereira, “Sensations, perceptions and emotions: Towards quality of experience evaluation for consumer electronics video adaptations,” in *Proceedings of First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.

[148] J. Radun, T. Virtanen, G. t. Nyman, and J.-L. Olives, “Explaining multivariate image quality–interpretation-based quality approach,” in *Proceedings of ICIS 06*, 2006, pp. 119–121.

[149] U. Reiter and T. Köhler, “Criteria for the subjective assessment of bimodal perception in interactive AV application systems,” in *IEEE/ISCE’05 International Symposium on Consumer Electronics*, 2005.

[150] D. M. Rouse, R. Pépion, P. Le Callet, and S. S. Hemami, “Tradeoffs in subjective testing methods for image and video quality assessment,” in *Proceedings of SPIE*, vol. 7527, 2010, pp. 75 270F–75 270F–11, online available at http://dx.doi.org/10.1117/12.845389, last seen on 22nd Nov. 2013.

[151] G. Lorho, “Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction,” in *Proceedings of Audio Engineering Society 119th Convention, New York (NY), USA*, vol. 6629, 2005.

[152] ——, “Perceived quality evaluation: An application to sound reproduction over headphones,” Ph.D. dissertation, Aalto University, School of Science and Technology, Espoo, Finland, 2010.

[153] M. Borrego, E. P. Douglas, and C. T. Amelink, "Quantitative, Qualitative, and Mixed Research Methods in Engineering Education," *Journal of Engineering Education*, pp. 53–66, 2009.

[154] H. Levitt, "Transformed Up-down Methods in Psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 467–477, 1970.

[155] G. T. Fechner, *Elemente der Psychophysik Teil 1.* Breitkopf und Härtel, 1860.

[156] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 2nd ed. Sage Publications, Thousand Oaks, California, USA, 2011.

[157] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "Open Profiling of Quality. A mixed method approach to understand multimodal quality perception," *Journal on Advances in Multimedia*, vol. 2010, no. 658980, p. 28, 2010, online available at http://dx.doi.org/10.1155/2010/658980, last seen on 22nd Nov. 2013.

[158] S. Bech, R. Hamberg, M. Nijenhuis, K. Teunissen, H. Looren de Jong, P. Houben, and S. K. Pramanik, "Rapid perceptual image description (RaPID) method," in *Proceedings of SPIE*, vol. 2657, 1996, pp. 317–328, online available at http://dx.doi.org/10.1117/12.238728, last seen on 22nd Nov. 2013.

[159] D. L. Morgan, *Focus group as qualitative research.* Sage, London, 1988.

[160] F. P. Robinson, *Effective Study. Revised Edition.* Harper & Brothers, Publishers, New York, 1961.

[161] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: strategies for qualitative research.* Aldine, Chicago, USA, 1967.

[162] A. Smolic, K. Mueller, P. Merkle, T. Rein, M. Kautzner, P. Eisert, and T. Wiegand, "Representation, coding, and rendering of 3D video objects with MPEG-4 and H.264/AVC," in *Multimedia Signal Processing, 2004 IEEE 6th Workshop on*, sept.-1 oct. 2004, pp. 379–382.

[163] C. Kuster, N. Ranieri, H. Zimmer, Agustina, J. Bazin, C. Sun, T. Popa, and M. Gross, "Towards Next Generation 3D Teleconferencing Systems," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, 2012.

[164] S. Nowak, “Evaluation Methodologies for Visual Information Retrieval and Annotation,” Ph.D. dissertation, Technische Universität Ilmenau, Germany, 2012.

[165] K. Kunze, D. Strohmeier, and S. Jumisko-Pyykkö, “Comparison of two mixed methods approaches for multimodal quality evaluations: Open Profiling of Quality and Conventional Profiling.” in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, 2011, pp. 137–142, online available at http://dx.doi.org/10.1109/QoMEX.2011.6065691, last seen on 22nd Nov. 2013.

[166] D. Embrey, *Task Analysis Techniques.* Human Reliability Associates Ltd., 2000.

[167] T. Alpert and J.-P. Evain, “Subjective quality evaluation - The SSCQE and DSCQE methodologies,” European Broadcastin Union(EBU), Tech. Rep., 1997, technical Review Spring.

[168] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, “User experience evaluation methods: current state and development needs,” in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ser. NordiCHI ’10. New York, NY, USA: ACM, 2010, pp. 521–530, online available at http://doi.acm.org/10.1145/1868914.1868973, last seen on 22nd Nov. 2013.

[169] J. Radun, T. Leisti, J. Häkkinen, H. Ojanen, J.-L. Olives, T. Vuori, and G. Nyman, “Content and quality: Interpretation-based estimation of image quality,” *ACM Trans. Appl. Percept.*, vol. 4, no. 4, pp. 2:1–2:15, feb 2008, online available at http://doi.acm.org/10.1145/1278760.1278762, last seen on 22nd Nov. 2013.

[170] N.-L. Hsueh, P.-H. Chu, P.-A. Hsiung, and M.-J. Chuang, “Supporting Design Enhancement by Pattern-based Transformation,” in *Proceedings of the 34th Annual IEEE Computer Software and Applications Conference*, 2010, pp. 462–467.

[171] S. Möller, K.-P. Engelbrecht, C. Kühnel, , I. Wechsung, and B. Weiss, “A Taxonomy of Quality of Service and Quality of Experience of Multimodal Human-Machine Interaction,” in *Proceedings of the QoMEX 2009, First International Workshop on Quality of Multimedia Experience (July 19–31, San Diego, California, USA)*, 2009, pp. 7–12.

[172] D. Ramanan and S. Baker, "Local Distance Functions: A Taxonomy, New Algorithms, and an Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 794–806, 2011.

[173] C. Yu-Chun, C. Chi-Jui, C. Kuan-Ta, and L. Chin-Laung, "Radar chart: Scanning for high QoE in QoS dimensions," in *Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee on*, 2010, pp. 1–6.

[174] T. Hai Anh and A. Mellouk, "Real-Time state-dependent routing based on user perception," in *Communications and Information Technology (ICCIT), 2011 International Conference on*, 2011, pp. 160–166.

[175] C. Mangtang and L. Kecheng, "Modeling Media Synchronization with Semiotic Agents," *MultiMedia, IEEE*, vol. 17, no. 1, pp. 14–23, 2010.

[176] D. Chaoqun, "Study on application of semiotics theories in identifier design," in *Computer-Aided Industrial Design Conceptual Design, 2009. CAID CD 2009. IEEE 10th International Conference on*, 2009, pp. 507–511.

[177] L. Bento, R. Prates, and L. Chaimowicz, "Using Semiotic Inspection Method to Evaluate a Human-Robot Interface," in *Web Congress, 2009. LA-WEB '09. Latin American*, 2009, pp. 77–84.

[178] A. Mittal, A. K. Moorthy, J. Ghosh, and A. C. Bovik, "Algorithmic assessment of 3D quality of experience for images and videos," in *IEEE Digital Signal Processing Workshop*, 2011.

[179] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 13, no. 5, pp. 882–893, October 2011, online available at http://dx.doi.org/10.1109/TMM.2011.2157333, last seen on 22nd Nov. 2013.

[180] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless Video Quality Assessment: A Study of Subjective Scores and Objective Algorithms," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 4, pp. 587–599, 2010.

[181] D. Cerra, A. Mallet, L. Gueguen, and M. Datcu, "Algorithmic Information Theory-Based Analysis of Earth Observation Images: An Assessment," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 1, pp. 8–12, 2010.

[182] S. Cardeal, F. Neves, S. Soares, F. Tavares, and P. Assuncao, "ArQoS: System to monitor QoS/QoE in VoIP," in *EUROCON - International Conference on Computer as a Tool (EUROCON), 2011 IEEE*, 2011, pp. 1–2.

[183] H. Du, C. Guo, Y. Liu, and Y. Liu, "Research on relationship between QoE and QoS based on BP Neural Network," in *Network Infrastructure and Digital Content, 2009. IC-NIDC 2009. IEEE International Conference on*, 2009, pp. 312–315.

[184] M. Gershon, "A perspective on quality of experience and quality of service parameters in IPTV," in *Microwaves, Communications, Antennas and Electronics Systems, 2009. COMCAS 2009. IEEE International Conference on*, 2009, pp. 1–4.

[185] K. Hyun-Jong and C. Seong-Gon, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," in *Advanced Communication Technology (ICACT), 2010 The 12th International Conference on*, vol. 2, 2010, pp. 1377–1382.

[186] B. Wang, X. Wen, S. Yong, and Z. Wei, "A New Approach Measuring Users' QoE in the IPTV," in *Circuits, Communications and Systems, 2009. PACCS '09. Pacific-Asia Conference on*, 2009, pp. 453–456.

[187] K. Hyun-Jong and C. Seong-Gon, "Design of QCS with QoE parameters on multimedia service and its performance analysis," in *Optical Internet (COIN), 2010 9th International Conference on*, 2010, pp. 1–3.

[188] R. Lopes Gomes, J. Jailton, W. Moreira, and A. J. Abelém, "QoE and QoS in wireless mesh networks," in *Communications, 2009. LATINCOM '09. IEEE Latin-American Conference on*, 2009, pp. 1–6.

[189] L. Yebin, D. Qionghai, and X. Wenli, "A Point-Cloud-Based Multiview Stereo Algorithm for Free-Viewpoint Video," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 3, pp. 407–418, 2010.

[190] E. Ekmekcioglu, M. Mrak, S. Worrall, and A. Kondoz, “Edge adaptive upsampling of depth map videos for enhanced free-viewpoint video quality,” *Electronics Letters*, vol. 45, no. 7, pp. 353–354, 2009.

[191] F. Bruls, S. Zinger, and L. Do, “Multi-view coding and view synthesis for 3DTV,” in *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, 2011, pp. 685–686.

[192] A. Douglas, R. Walters, and G. Wills, “NPE - A Conceptual Model and Language for the Representation of Norms,” in *Quality of Service-Based Systems (in conjustion with 10th International IEEE Conference on Quality Software)*, April 2010, event Dates: 14/15 July 2010. Online available at http://eprints.soton.ac.uk/270940/.

[193] G. Cheng, K. Raahemifar, and O. Das, “Comparative study of heuristics for reliability optimization of complex systems,” in *Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on*, 2010, pp. 1–4.

[194] U. . QuestBack, “Unipark,” product website, online available at www.unipark.info, last seen on 19th September 2016, 2014.

[195] F. ten Kleij and P. A. Musters, “Text analysis of open-ended survey responses: a complementary method to preference mapping,” *Food Quality and Preference*, vol. 14, no. 1, pp. 43–52, 2003, online available at http://www.sciencedirect.com/science/article/pii/S0950329302000113, last seen on 22nd Nov. 2013.

[196] J. Fahrenberg and H. Selg, *Das Freiburger Persönlichkeitsinventar FPI, Handanweisung für die Durchführung und Auswertung.* Verlag für Psychologie, Dr. C. J. Hogrefe, Göttingen, GER, 1970.

[197] J. M. Carroll and J. Reitman Olson, *Mental models in human computer interaction.* Washington, D.C.: National Academic Press., 1987.

[198] Datacolor, “Display Calibration,” product website, online available at http://spyder.datacolor.com/display-calibration/, last seen on 19th September 2016, 2016.

[199] I. The MathWorks, “MATLAB - The Language of Technical Computing,” product website, online available at http://www.mathworks.de/products/matlab/, last seen on 19th September 2016, 1994–2016.

[200] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/

[201] S. Jumisko-Pyykkö, U. Reiter, and C. Weigel, "Produced Quality is Not Perceived Quality - A Qualitative Approach to Overall Audiovisual Quality," in *3DTV Conference, 2007*, may 2007, pp. 1–4.

[202] F. Husson, S. Lê, and J. Pagès, "Confidence ellipse for the sensory profiles obtained by principal component analysis," *Food Quality and Preference*, vol. 16, no. 3, pp. 245–250, 2005, online available at http://www.sciencedirect.com/science/article/pii/S0950329304000631, last seen on 22nd Nov. 2013.

[203] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, pp. 33–51, 1975, online available at http://dx.doi.org/10.1007/BF02291478, last seen on 22nd Nov. 2013.

[204] M. Shahid, A. Rossholm, and B. Lövström, "A No-Reference Machine Learning Based Video Quality Predictor," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX 2013)*, 2013, pp. 176–181.

[205] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience," *Telecommunication Systems*, vol. 52, no. 2, pp. 587–600, 2013, online available at http://dx.doi.org/10.1007/s11235-011-9503-7, last seen on 22nd Nov. 2013.

[206] O. U. Press, "recommend," Online available at http://www.oxforddictionaries.com/definition/english/recommend, last seen on 3rd Dec. 2013, 2013.

[207] J. Nielsen, *Usability Engineering.* Morgan Kaufmann, San Francisco, 1993.

[208] Bibliographisches Institut & F. A. Brockhaus AG, "Brockhaus Enzyklopädie - Ontologie," Online available at http://www.brockhaus-enzyklopaedie.de/be21_article.php#1, last seen on 22nd Nov. 2013, 2005–2013.

[209] H. Moller, "Multi-dimensional audio," Leonardo Mani Audio, Tech. Rep., 1977, also available at http://www.linkwitzlab.com/multidimensionalaudio.pdf, last seen on 22nd Nov. 2013.

[210] J. Baumeister, “Steuerungstheorie,” lecture notes, in German language, online available at http://www.math.uni-frankfurt.de/~baumeist/s00-shell.pdf, last seen on 22nd Nov. 2013, 2000.

[211] A. Redl, C. Keimel, and K. Diepold, “Saliency Based Video Quality Prediciton Using Multy-Way Data Analysis,” in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX 2013)*, 2013, pp. 188–193.

[212] W. Saaty, R, “The Analytic Hierarchy Process - What it is and how it is used,” *Mathl Modelling*, vol. 9, no. 3–5, pp. 161–176, 1987.

[213] S. D. Amnon and F. Joyce E., “Quantifying Perceptual Image Quality,” in *PICS*, 1998, pp. 242–246.

[214] C. Mantel, S. r. Forchhammer, and J. Korhonen, “Reducing the duration fo paired-comparison experiments for visual quality assessment,” in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX 2013)*, 2013, pp. 50–51.

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
3DAV	3D Audio-Visual (MPEG group working on)
3DMC	3D mesh coding (static)
3DTV	Three-Dimensional Television
3DVO	Three-Dimensional Video Object
3DVQM	3D Video Object Quality Metric
4D	Four Dimensional
ACR	Absolute Category Rating
AFC	Alternative Forced Choice (also called Adaptive Forced Choice or IFC - Interval Forced Choice)
AFX	Animation Framework eXtension (MPEG-4 specification part)
AHP	Analytic Hierarchy Process
AOI	Areas of Interest
AR	Augmented Reality
ARCU	Application Resource Context User
BBC	British Broadcasting Channel
CAD	Computer-Aided Design
CGI	Computer-Generated Imagery
CVQE	Continuous Video Quality Evaluation
DCR	Degradation Category Rating
DFG	Deutsche Forschungsgemeinschaft (Engl.: German Research Foundation)
DIN	Deutsches Institut für Normung (Engl.: German Institute for Standardization)
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
DVD	Digital Video Disc

DVQ	Digital Video Quality
EBU	European Broadcasting Union
ec	Eye contact
FCP	Free Choice Profiling
FKTG	Fernseh- und Kinotechnische Gesellschaft (Engl.: German Association for TV and Cinema Technology)
FPI	Freiburger Persönlichkeitsinventar (Engl.: Freiburg Personality Inventory)
FUN	Fidelity, Usefulness, and Naturalness
FVV	Free Viewpoint Video
GPA	Generalized Procrustes Analysis
GT	Ground Truth
GUI	Graphical User Interface
hc	Horizontal correction
HCI	Human-Computer Interaction
HHI	Fraunhofer Heinrich Hertz Institute
HVS	Human Visual System
IBQ	Interpretation-based Quality
IDMT	Institute for Digital Media Technology
IPA	Interpretive Phenomenological Analyis
ISO	International Organization for Standardization
ITU	International Telecommunication Union
I3D	incomplete 3D
JND	Just Noticeable Difference
LCD	Liquid Crystal Display
LDI	Layered Depth Images
LQR	Low Quality Regime
MCCT	Matlab Camera Calibration Toolbox
MIQM	Multi-view Image Quality Measure
MOS	Mean Opinion Score
MPEG	Motion Picture Expert Group
MPQM	Moving Picture Quality Metric
MSE	Mean Square Error
MVC	Multi-view Video Coding
NVFM	Normalization Video Fidelity Metric
OPQ	Open Profiling of Quality

PC	Personal Computer
PCA	Principal Component Analysis
PoC	Proof of Concept
PSNR	Peak Signal-to-Noise Ratio
px	Pixel
QE	Quality Elements (service domain)
QF	Quality Features (perception domain)
QoE	Quality of Experience
QoS	Quality of Service
QUAL	qualitative
QUAN	quantitative
RaPID	RaPID perceptual image description method
SAMVIQ	Subjective Assessment Methodology for Video Quality
Skalalgo3d	Scalable algorithms for 3DVO under consideration of subjective quality factors
SNHC	Synthetic-Natural Hybrid Coding
SSCQE	Single Stimulus Continuous Quality Scale
SQ3R	Survey, Question, Read, Recite and Review
s3DVQM	Subjective 3D Video Object Quality Metric
TA	Task analysis
TR	Technical Realization
TV	Television
UGC	User Generated Content
USB	Universal Serial Bus
UX	User Experience
VDM	Visual Differences Predictor
VQEG	Video Quality Expert Group
VQM	Video Quality Metric
ZDF	Zweites Deutsches Fernsehen (Engl.: 2nd channel of German TV broadcasting)

List of Figures

List of Tables

Appendices

Appendix A - Evaluation Material, Additional Analysis Results

This section contains additional information gathered from all the evaluation activities. The structure follows the structure of the thesis. First of all, additional material from the explorative studies is provided that includes the interview guideline for the expert interviews, additional information about the focus groups, and the online questionnaire. Second, the pre-test information for study 1, the laboratory experiment, can be found here. Third, additional information concerning study 2, the OPQ experiment, is given.

Expert Interviews:
The translated interview guideline is presented below; the interview was conducted in German.

Introduction:
One aim of my work is to contribute to quality improvement of 3D video. The survey is part of the project "Scalable algorithms for 3D video objects under consideration of subjective quality factors". This project is conducted at the Institute of Media Technology of the Technische Universität Ilmenau (Thuringia, Germany) and is supported by the German Research Foundation (short: DFG). In general, it is about finding a good way to assess 3D video quality that takes into account the potential usage (users' expectations, users' pre-experience with the handling of 3D video) and the individual quality assessment of every single user. I ask you to help me and to share with me your insights, your experiences, and your knowledge by answering the following questions. This will take approximately 15 minutes of your time. Your privacy is a great concern for me and, thus, all your statements will be treated anonymously at every stage of data usage and discussed confidentially in my research work. Other survey participants or people from outside do not have access to information about other participants' answers or data. After the analysis, evaluation, and completion of the research project,

all your personal data will be destroyed. Please do not hesitate to contact me via e-mail (Sara.Kepplinger@tu-ilmenau.de), if there are still open questions. Thank you very much for your participation!

1. Pre-Experience:
What is / was your motivation to work on the topic of 3D video (objects)?

1.1 In which area(s) have you mostly worked?

- 3D video object creation: Please mention the used procedure: (Picture-based? Model-based?)
- Free view point video (in general)
- Silhouettes (e.g., shape from silhouette procedure)
- Coding / Compression procedures
- Standardization
- (objective) Quality assessment of 3D video
- Synthesis, namely:
- Others:

1.2 Which role did the topic of quality play in this case / these cases? Please describe:

2. Questions on picture quality: How do you estimate good quality of a video picture for yourself?

2.1 How do you define the concept of quality in connection with a 3D video object?

2.2 What are typical quality factors of 3D video objects for you? Please mention at least three:

2.3 If you compare 2D with 3D video: Where do you notice differences related to quality factors?

2.4 In your opinion, to which extent does interaction influence the quality assessment of the picture during the viewing of a 3D video (object)? Please explain:

2.5 In your opinion, which characteristics of 3D videos or 3D video objects are relevant, at the moment, for current typical and critical testing material, and where do you see big changes in the future?

Objective measurement of quality usually takes place with typical or critical testing material, which should include representative test sequences. The usual material or the material defined as usual is adjusted again and again to the actual status of technology and to the defined standard.

2.6 How or with which means would you measure the quality of a 3D video / video object?

- 2.6.1 How would you proceed?
- 2.6.2 Why?

2.7 Which procedure for 3D video (object) creation do you think has chances in the future if it provides good enough quality to the user? Please explain briefly:

3. Questions on the actual usage of 3D: Do you also use 3D in your spare time? Please identify which of these categories are relevant for you:

- Yes, 3D cinema with glasses
- Yes, 3D TV with glasses
- Yes, 3D TV without glasses
- Yes, 3D videos on www.YouTube.com with glasses
- Yes, 3D video conferences on the PC
- Yes, 3D computer games
- Yes, play station with 3D video on TV devices
- Yes, mobile 3DTV on the cell phone
- Yes, 3D architecture or other software programs for technical drawings (e.g., Maya. . .)
- Yes, 3D displays of sights (e.g., virtual walk in a museum)

- Yes, I have once been to a so-called CAVE (= Cave Automatic Virtual Environment) (e.g., at the ARS Electronica, at a University. . .) or in a similar installation, namely:
- Yes, I have already made 3D recordings on my own.
- No, I am just concerned with this topic due to my work. Because:
- No, I am 3D blind.
- Others:

3.1 In your opinion, what would be the ultimate, so-called killer application using 3D videos (video objects) and which characteristics should this application have? Please describe: 3.2 Which major need(s) for optimizing 3D video (objects) do you see? Please illustrate:

4. Demographic Questions: sex:

- male
- female

4.1 Age: Please indicate your year of birth:

- before 1934
- between 1934 and 1949
- between 1949 and 1964
- between 1964 and 1979
- between 1979 and 1994
- between 1994 and 2000

4.2 Education: Please indicate what is your highest educational qualification:

- no graduation
- basic education
- completed vocational training

- junior high school
- other graduation
- High School Certificate (Abitur)
- University degree (Bachelor, Master, Diploma…)
- doctorate

4.3 Household / Income: Please indicate your monthly income for the entire household (=gross total income of all people living in one household):

- less than 1000 Euros
- between 1000 and 3500 Euros
- between 3500 and 5000 Euros
- between 5000 and 6500 Euros
- between 6500 and 8000 Euros
- more than 8000 Euros

4.4 Family status: Please indicate your family status (multiple answers possible):

- single
- living with a partner
- in a relationship with separate households
- married
- father/mother of one child
- father/mother of 2 children
- father/mother of more than 2 children

Thank you very much!

The relevant analysis results in this thesis are presented in Chapter 4.2.1.

Focus Group Material:
The following guideline document, including the questions asked, was provided in German. Here, it is presented in English translation.

Focus Group:
The intention is to analyze expectations, wishes, and ideas concerning 3DVOB, in general, and free viewpoint choice, in particular. Maximally 8 participants per focus group receive a preview and the possibility to watch prototypical test sequences of 3DVOs.

Aim:
Presentation of a possible use case (video communication system) to pre-experienced and inexperienced users, in order to collect and evaluate wishes, ideas, and expectations concerning the usage of 3DVOs and free viewpoint selection.

Material to be prepared beforehand:

1. Note questions which have to be answered during the focus group meeting.
2. Record test sequences and prepare them for the focus group.
3. Prepare data privacy statement, questionnaire, flash cards, posters, 8 pens, glue, and cellotape.
4. Provide a descriptive documentation about the functionality of the system.
5. Prepare the flash card technique for the evaluation of the usage of video communication in everyday life.
6. Prepare a questionnaire for demographic data and pre-experience.
7. Prepare the request for a down payment for the expenditures for the test (payment for max. 16 participants)
8. Recruit participants (run an advertisement, send e-mails, mail- and telephone answering service).
9. Organize and reserve room (incl. preparation time before and afterwards).
10. Organize recording device (mini DV cam, dictaphone) and presentation equipment (laptop, projector, files of video sequences).

Checklist prior to focus group meeting:

- Room prepared
- Data privacy statement, questionnaire, flash cards, posters, 8 pens, glue, cellotape
- List of participants, mobile phone numbers
- Recording devices ready
- Presentation equipment ready (2x cams, power, tripods, projector, laptop, videos)
- Drinks and breadsticks
- Payment and confirmation of money and receipt, contact list for future events

Procedure:
The following Tables 6.1 and 6.2 give an example of a guideline for the test instructor. This includes the duration of each activity as well as a distinction between the test participants' and the instructors' role.

Table 6.1: Focus group procedure used.

Participants	Instructor	Duration
Participants enter the room.	Welcome by the instructor, hand out of demographic questionnaire, allocation of space	00:05:00
Participants sit down and listen.	Welcome address and introduction:“Good afternoon! I am happy that you all are present today to participate in the focus group. Today the goal is to give you insights into the research activities at the TU Ilmenau and, at the same time, to ask you as experts for this.”	00:10:00
Participants sign data privacy statement	Instructor hands out forms for the signature:“As I am making audiovisual records for later analysis, and a statement you make or picture (without names) of you might be used for research publications, I would like to ask you for your acceptance of the data privacy statement. During this focus group meeting I will show you a not yet published system or system idea. Therefore, I ask you with this data privacy statement also to keep the information you get private. This statement covers the TU Ilmenau as well as your privacy and has to be signed before we start with the focus group.”	00:15:00
Participants listen and note the presentation of the system.	Instructor: Presentation of the project and the system with descriptive documentation video. “Skalalgo3d aims at the development of a video communication system that provides the possibility to the user to create eye contact with the conversation partner without much effort. Perhaps, the system will provide more than that: namely, to allow the user a free viewpoint choice. . . Until now it is not possible to have direct eye contact and simultaneously observe the communication partner because one has either to look into the camera (to create the feeling of eye contact), or one looks at the displayed conversation partner (and therefore is not able to look into the camera). Skalalgo3d uses the possibility to create virtual views, for which an additional perspective is created out of several camera views (recordings). This may lead to a simulated eye contact. In this case, it is possible to separate foreground from background information and, with this technological possibility, users of the video communication system could also choose their own preferred viewpoint of the scene. (Instructor shows videos for further explanation.) My question for you is now, if you could imagine using such an additional application, and what you would use it for, and which advantages and disadvantages you see. . . ”	00:35:00
Participants work on the tasks.	Task round: Instructor gives tasks to the test participants that should be solved with a practical, apparent and prototypical usage of the presented system. “You will now get four tasks from me, which you should solve one after another. Please make a note, during these activities, as soon as you have an idea for improvement, or you recognize something, or you have a vision for the future, or something similar. Of course, you are the experts and you should also make a note of problems and discrepancies.”	00:36:00
	Task 1 - Warm up	(00:07:00)
	Instructor: hands out test material (flash cards) to the participants. “First of all, I would like to hear from you about your current video communication usage in your daily routine. Do you use video communication? When? With whom do you communicate via video communication system (only close friends, relatives, business partner, or with everybody)? Why are you communicating via a video communication system? Is the video important for a special reason? Is it simply the usage of video communication or does this also happen sometimes alongside other activities? I have prepared some cards for you and a few small "‘everyday life"’ posters. Please position your cards in such a way that I am able to get a picture about your video communication usage. For this, you will have five minutes.”	00:43:00
	Task 2 - Pairwise video communication simulation	(00:05:00)
	“Please get together in pairs and pretend that you are communicating via video with each other. Please make notes if something attracts your attention. Which factors are important during such communication? What should not be missing? What would be nice, if it were available - additionally? Please do not forget that at the moment, eye contact is not possible.”	00:48:00

Table 6.2: Focus group procedure used (continued).

Participants	Instructor	Duration
	Task 3 - Requirement definition	(00:05:00)
	“You reconstructed a video communication situation. . . Now - in 5 minutes: What are your requirements for a video communication system? Can you imagine other situations or a usage in another environment? Please be visionary and think about situations and examples under the usage of a video communication system that would be completely new to you but incredible great and useful for you. Note all important details: you will have the opportunity in the final discussion round to explain your ideas and show why they are important.”	00:53:00
	Task 4 - Final task/Stimulation for the discussion	(00:07:00)
	“Now take your "‘everyday life"’ poster again. You described, in the first task, how, when, and where you use video communication at the moment. Please expand your poster now by including the knowledge you collected during the last half hour, your visions and ideas, and, with your knowledge about the Skalalgo3d project, how your video communication usage in the daily routine would change (e.g., because of new usage possibilities): Where do you see changes? What would stay the same? Are there any disadvantages? What are the advantages? Which influence would eye gaze correction have for your video communication usage? Please note everything that comes to mind. You will also be able to include this in the upcoming discussion round.”	01:00:00
Participants answer and ask questions and discuss	Discusson, Instructor moderates the discussion, asks questions. “Let’s now come to a discussion round that will take 20 minutes. You got to know the system and the concept behind through the presentation and the separate tasks and you now may be in position to answer some of the following questions. My first question to you, as a user is, (use following trigger questions, open and without obligation): 1. Which advantages and disadvantages do you see for this presented system? 2. How should a perfect video that provides free viewpoint choice and eye gaze correction function? 3. Does the decision about the camera setting and the number of cameras in use play a role for the user? 4. Up to which level can one ask for activity by the user for the calibration? What is still acceptable? 5. Did you notice a better video quality as soon as a viewpoint correction was made? 6. How do you perceive the difference between the real background and the blue screen? 7. How would you imagine the usage of the system with more than two participants or in a group?”	01:30:00
Wrap up	Instructor collects test material from the test participants.	01:35:00
Participants sign confirmation about receipt of money, and possibly enroll in list for future participation	Payment Instructor collects confirmations about receipt and hands out the payments. “Thank you very much for your participation! As agreed, everyone of you gets an expense allowance. Therefore I ask you to sign the confirmation that you have received the money. If you would like to participate in future evaluation activities, please, enroll in this contact list.”	01:50:00
Participants leave the room.	Farewell, Instructor says good bye	02:00:00

Checklist for afterwards:

- Data privacy statement filed
- Confirmation of money receipt filed
- List of participants and expenditures complete, for accounting
- Recordings stored, archived, and transcribed
- Recording devices and presentation equipment returned
- Leftover of drinks and breadsticks disposed of
- Leave room in initial state
- Accounting

The relevant analysis results in this thesis and related to this section are presented in Chapter 4.2.1.

Online Questionnaire:
The following text has been used in the online questionnaire (in German, the following text is translated); the results are presented, in part, in Chapter 4.2.1:

Your opinion is wanted:
"Videoquality, 3D Quality Factors and the User's Wishes" within the research project "Scalable algorithms for 3D video objects under consideration of subjective quality factors"

Introduction:
One aim of my work is to contribute to the quality improvement of 3D video. The survey takes place within the project "Scalable algorithms for 3D video objects under consideration of subjective quality factors". This project is conducted at the Institute of Media Technology of the Technische Universität Ilmenau (Thuringia, Germany) and is supported by the German Research Foundation (short: DFG). In general, it is about finding a good way to assess 3D video quality that takes into account the potential usage (users' expectations, users' pre-experience with the handling of 3D video) and the individual quality assessment of every single user. I ask you to help me and to share with me your insights, your experiences, and your knowledge by answering the following questions. This will take approximately 15 minutes of your time. Your privacy is an important concern for me and all your statements will be treated anonymously at every stage of data usage and discussed confidentially in my research work. Other survey participants or people from outside do not have access to information about other participants' answers or data. After the analysis, evaluation, and completion ofthe research project, all your personal data will be deleted. However, please consider that, because of the usage of Unipark, I cannot guarantee the same confidential treatment of data at Unipark's activities. Please do not hesitate to contact me via e-mail (Sara.Kepplinger@tu-ilmenau.de), if there are still open questions.
Thank you very much for your participation!

Participation Agreement: Do you agree with moving on? Please indicate:

- I have read the introduction and I wish to continue.
- No, I do not want to proceed.

1. *Pre-experience:* Do you have experience with 3D? Please indicate::
   - Yes, 3D cinema with glasses

- Yes, 3D TV with glasses
- Yes, 3D TV without glasses
- Yes, 3D videos on www.YouTube.com with glasses
- Yes, 3D video conference on the PC
- Yes, 3D computer games
- Yes, play station with 3D video on TV devices
- Yes, mobile 3DTV on a cell phone or another mobile device
- Yes, 3D architecture or other software programs for technical drawings (e.g., Maya...)
- Yes, 3D displays of sights (e.g., virtual walk in a museum)
- Yes, I have once been to a so-called CAVE (= Cave Automatic Virtual Environment) (e.g., at the ARS Electronica, at a University)
- or in a similar installation, namely:
- Yes, I have already made 3D recordings on my own.
- No, I am just concerned with this topic due to my work. Because:
- No, I am 3D blind.
- Others:

1.1 What is / was your motivation to work on the topic of 3D video (objects)? Please describe your experiences here. Or write about the reasons why you do not have experience with 3D video.

1. Usage: You have the exclusive chance to watch a movie in 3D. Now you can choose: Which genre would you take because of the 3D display option? Why?

- I don't care about the genre. Because:
- Crime. Because:
- Documentary. Because:
- Action. Because:

- Romance. Because:
- Science-Fiction. Because:
- Comedy. Because:
- Western. Because:
- Drama. Because:
- Fantasy. Because:
- Adventure. Because:
- Animation. Because:
- Others, for example: . . . Because:

2.1 Interaction: Imagine you are watching a 3D VIDEO and you have two options. Which option would you choose? Please indicate:

- I would like to decide on my own from which side / perspective I can watch the image.
- I would like to have predefined views displayed, where I do not have to decide on a direction of view.

Please describe the reason for your choice:

2.2 Idea of benefit by 3D technique: In which situation would you benefit most from a 3D video representation? Please describe:

3. Questions on picture quality: In general, how do you estimate the quality of a video picture?

3.1. How do you define the concept of quality in connection with a 3D video object?

3.2. What are typical quality factors of 3D video objects for you? Please note at least three:

3.3. If you compare 2D with 3D: Where do you notice differences concerning quality factors?

3.4 How, or on the basis of which criteria, would you judge the quality of 3D? (How would you proceed in this case? Why?) Please give a brief description:

3.5. In your opinion, which 3 advantages and 3 disadvantages would 3D video offer you? Please insert:

4. Demographic Questions:

- male
- female

4.1. Age: Please indicate your year of birth:

- before 1934
- between 1934 and 1949
- between 1949 and 1964
- between 1964 and 1979
- between 1979 and 1994
- between 1994 and 2000

4.2. Education: Please indicate your highest educational qualification:

- no graduation
- basic education
- completed vocational training
- junior high school
- other graduation
- High School Certificate (Abitur)
- University degree (Bachelor, Master, Diploma...)
- doctorate

4.3. Household / Income: Please indicate your estimated monthly income for the whole household (= gross total income of all people living in one household):

- too little
- not much
- enough
- a little bit more than enough
- a lot
- one could say too much

4.4 Family status: Please indicate your family status (multiple answers possible):

- single
- living with a partner
- in a relationship with separate households
- married
- father/mother of one child
- father/mother of 2 children
- father/mother of more than 2 children

5. Further Participation:

- Yes, I'd like to participate again in surveys and evaluation tests on 3D and try out new developments. My e-mail address to contact me on this issue is:
- No, please do not contact me again.

Thank you for your participation!

The relevant analysis results in this thesis and related to this section are presented in Chapter 4.2.1.

Laboratory Experiment:
In preparation for the laboratory experiment, a pre-experiment took place in order to define a possible influence of the shown content (i.e., different persons). Therefore, test participants were asked to answer an online questionnaire with questions about pictures of persons. These questions are based on the theory that the different persons shown are perceived differently (with respect to factors like trustworthiness and attractiveness) and that this may influence the overall quality rating of the shown test stimuli. The displayed persons can be seen in the following Figure 6.1. The pictures in the top row of Figure 6.1 exhibit persons with eye contact, while each person in the bottom row is depicted without eye contact. All the persons were recorded with the same background, recording situation (i.e., light setting, etc.), and the same T-shirt.

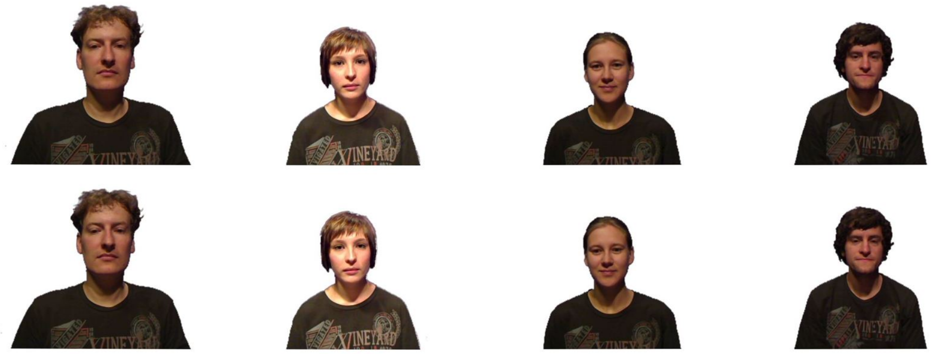


Figure 6.1: Pictures shown in the pretest.

The questionnaire started with a short introduction explaining the questionnaire and asking for 10 minutes of time to answer the following questions (translated, because the questionnaire was in German):

What is true for this person? (7-point scale labeled with “it applies” on the left and “it is not the case” on the right)

- attractive
- inspires confidence
- pleasant
- credible

- honest

Please tell us your sex:

- male
- female

Please tell us your age:

Thank you for your participation!

The results of this experiment constitute a collection of 41 data sets. The analysis (T-test for related samples computed with SPSS) was carried out once for all five attributes (i.e., attractive, inspiring confidence, etc.) with the ratings for the test stimuli with eye contact and once for the stimuli without eye contact. The results show that there are no significant differences between the different contents. However, comparing the summarized ratings of stimuli with and without eye contact leads to a result of significant differences concerning the attributes *inspiring confidence* (t -2.57) and *credible* (t -2.64) and to a remarkable difference concerning the attribute *honest* (t -.93). This may lead to the hypothesis that eye contact becomes important concerning aspects like honesty and that this is true for every shown person, whereas persons shown differently do not significantly influence such aspects.

OPQ:
In the following Table 6.3, the definitions of described/clustered attributes are provided in German and English. The experiment was conducted in German. Because the individual attributes are perceived subjectively by the test participants, it is important to ensure a consensus on the original definitions.

Table 6.3: Attributes generated in German and translated into English.

GERMAN	ENGLISH	GERMAN	ENGLISH
flimmern, flimmernd	flicker	Störung	disturbance
unruhig	uneasy, shaky	gezackt	jagged
Blick Richtung Kamera	view into camera	verpixelt, pixelig, Pixelfehler, Verpixelung	pixelated
Ränder unscharf, verwackelte Ränder	fuzzy edges	unscharf, Unschärfe	blurred
wave im Gesicht	wave in the face	unnatürlich	unnatural
flatternd	flittering	verschwommen, verwischt	blurry, blurred
fremd	strange	verzerrtes Gesicht, Verzerrung im Gesicht (entstellte Gesichtszüge)	deformed face, deformations in the face (ghastly, deformed facial expressions)
geteilt	parted	unförmig	misshapen
langsam	slow	falsche Bildinhalte	wrong picture contents
Farbfehler, Farbstörung	wrong color	Auflösung	resolution
ausgeschnitten	cut out	randlos	frameless
verzogen	distorted	Farbflecken	color spots
gemischte Farbübergänge	mixed color transitions	sehr abgehackt	very choppy

Table 6.3: Attributes generated in German and translated into English.(continued)

GERMAN	ENGLISH	GERMAN	ENGLISH
Arm überdeckt	overlaid arm	verwaschene Kanten (Höhenkanten-artefakte)	blurred edges
hell	bright	vollständig bzw. unvollständig	complete or not complete, incomplete
Vollständigkeit des Objektes	completeness of the object	Abgrenzung zum Hintergrund, scharfe Abgrenzung des Hintergrundes	differentiation to background, sharp differentiation to the background
wabbern und ausfransen	billow and fringe	Genauigkeit des Ausschneidens	precision of cutting out
Augenpartie gut dargestellt	good presentation of eye area	formtreu	formatwise conform
Mund versetzt, schief	mouth displaced	gleichmäßige Farbverläufe	consistent color gradient, steady color gradients
graue Flecken	grey spots	fehlerhaft	defective
transparent, durchsichtig	transparent	Loch, löchrig, Löcher im Bild, fehlende Bildinhalte, fehlende Teile, fehlend	hole, holey, holes in the picture, missing picture parts, missing parts, missing
Fragmente	fragments	Artefakte	artefacts
flackern	flicker	Rauschen, verrauscht	noise/noisy
zitternd	jittery	Verzerrungen in der Umgebung	deformations in the surrounding
nervös	nervous	Punkte	dots
ausgefranst	fringed	Bildausfall Kopf	picture breakdown head
flirrendes Bild Rumpf	shimmering picture at torso	Ameisen	ants

Table 6.3: Attributes generated in German and translated into English.(continued)

GERMAN	ENGLISH	GERMAN	ENGLISH
gruselig	creepy	Kanten flimmern	flickery edges
Konturen flimmern, flackernde Kontouren	flickery contours	Konturen verfälscht	distorted contours
verzerrt	contorted	wellig verzeichnet, Wellenartefakte	uneven distortions, wave artifacts
Linien im Bild, Linienbildung	lines in the picture, line creations	Bilder wo eigentlich nichts ist	pictures where usually there is nothing
Bildausfall Rumpf	picture breakdown at torso	Gleichmäßigkeit	consistency, equability
Blockartefakte	blocky	bewegungsanfällig	prone to movement
Streifen ziehen Kopf	stripes drawn at head	Schatten durch Punkte	shades through/by points
Spinnweben	spider's webs	grieseln, körnig, grießig	grainy
Komprimierungs-artefakte	compression artifacts	gestreckt	stretched
Schattierung	shading	pfleckig	spotted
Detailtreue	attention to detail	verwackelt an Haaren	fuzzy at hair parts
Streifen ziehen Rumpf	stripes drawn at torso	auslaufender Rand	edges flowing out
streifig	stripy	Schärfe	sharpness
störende Pixellinien	disturbing pixel lines		

Appendix B - Interlinking Model

Table 6.4: Numbers of mentioned QF per QE.

QF	incomplete, holey	blurred	contorted	stripes and lines	deformations, shape distortions	pixelated	flicker, jitter	color distortions, bad color quality	unlikely background	wrong proportions	slow, lagged	color spots, blemished
A1-S1	162	120	132	0	12	94	12	16	9	15	112	92
A1-S2	203	122	118	0	16	53	40	9	11	27	114	52
A2-S1	173	110	147	0	9	63	23	12	13	17	131	71
A2-S2	215	174	234	0	30	67	12	3	10	26	126	116
A3-S1	369	202	194	0	41	124	29	14	39	35	156	70
A3-S2	275	169	184	0	28	101	30	6	8	13	119	12

Table 6.5: Numbers of available QF in correlation circles (considering the determination coefficient) per QE.

QF	incomplete, holey	blurred	contorted	stripes and lines	deformations, shape distortions	pixelated	flicker, jitter	color distortions, bad color quality	unlikely background	wrong proportions	slow, lagged	color spots, blemished
A2-S3	15	10	0	1	13	4	14	5	1	2	2	4
A4-S4	0	9	9	6	13	6	3	3	0	3	2	4
A4-S5	1	6	7	8	2	4	8	2	0	3	2	3
A5-S4	2	7	2	7	10	8	9	4	1	10	1	6
A5-S5	4	7	5	10	2	7	11	1	1	4	1	2