

I n t e r n a t i o n a l T e l e c o m m u n i c a t i o n U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.916

(03/2016)

SERIES P: TERMINALS AND SUBJECTIVE AND
OBJECTIVE ASSESSMENT METHODS

Audiovisual quality in multimedia services

Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video

Recommendation ITU-T P.916

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than voice services	Series	P.1500

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.916

Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video

Summary

Recommendation ITU-T P.916 addresses issues related to assessing and minimizing visual discomfort and visual fatigue from 3D video. Issues addressed include 3D video characteristics that cause visual discomfort; symptoms of visual fatigue; relationships among the camera positions, the captured content and the viewer's perception; footage guidelines; viewer guidelines; information on the 3D; and viewing environment considerations.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.916	2016-03-15	9	11.1002/1000/11830-en

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2016

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

Page

1	Scope.....	1
1.1	Limitations.....	1
2	References.....	1
3	Definitions	1
3.1	Terms defined elsewhere	1
3.2	Terms defined in this Recommendation.....	1
4	Abbreviations and acronyms	3
5	Conventions	3
6	Technological issues for 3D displaying.....	3
6.1	Accommodation-vergence conflict	3
6.2	3D geometric distortions	4
6.3	Visual comfort zone or comfortable viewing zone	6
7	Binocular disparity relationship in capturing, storing and rendering	7
7.1	Conversion equations	7
7.2	Conversion of pixel disparities to a perceptual space	8
7.3	Typical constraints for displays.....	11
8	3D video content guidelines	14
8.1	Introduction	14
8.2	Disparity constraints.....	14
8.3	View asymmetries	16
8.4	Scene cuts	17
8.5	Excessive disparity/parallax	17
9	3D viewer guidelines	17
9.1	The influence of stereopsis and abnormal binocular vision (ABV).....	17
9.2	Warnings for subjects	18
9.3	Symptoms of 3D visual fatigue	18
10	3D viewing environment and playback guidelines.....	18
10.1	Relationship between viewer and 3D monitor	19
10.2	Viewing time and visual fatigue.....	20
11	Assessment methods for symptoms of visual fatigue.....	20
11.1	Questionnaire to assess symptoms of visual fatigue	20
11.2	Pair comparison (PC) subjective test to visual discomfort.....	20
	Appendix I – Open questions.....	22
	Appendix II – Visual fatigue questionnaire	23
	Bibliography.....	25



Documents

Rec. ITU-T P.916 (03/2016)

iii

Document Sharing Hub

Introduction

Three dimensional television (3DTV) systems are based on binocular disparity. Compared to the real world scenario, the reproduction of depth in 3DTV systems is limited by the camera position, number of cameras, source content and the technical limitations of the display technique. These include resolution, frame rate and view separation conditions. Current display technologies may also induce a conflict between accommodation (eyes focusing on the screen) and the convergence (eyes converging to a point in front of or behind the screen). These limitations can cause visual discomfort and visual fatigue.

Rec. ITU-T P.916 (03/2016)

Recommendation ITU-T P.916

Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video

1 Scope

This Recommendation includes information and guidelines for assessing and avoiding visual discomfort and visual fatigue from 3D video. It describes potential causes of visual discomfort and symptoms of 3D visual fatigue, including problems caused by watching stereoscopic and autostereoscopic television. This information is primarily intended to be used for the design and conduct of subjective assessment of the video or audiovisual quality of 3D video. This information is also appropriate for all 3D viewing experiences.

1.1 Limitations

This Recommendation includes guidance on recognizing visual fatigue symptoms and identifying 3D programming clip characteristics that may cause visual fatigue.

This Recommendation excludes objective assessment methods of 3D visual fatigue, because these are not computational algorithms but rather correspond to a medical diagnostic process performed by medical doctors.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T P.913] Recommendation ITU-T P.913 (2016), *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 accommodation-vergence conflict: Accommodation (eyes focussing on an object) and convergence (eyes pointing toward the object) are normally linked. Accommodation-vergence conflict is a conflict between accommodation (eyes focusing on the screen) and the convergence (eyes converging to a point in front of or behind the screen). Current 3D display technologies may induce accommodation-vergence conflict.

3.2.2 binocular disparity: The two eyes view an object from two different positions and angles. Due to the parallax, the object is projected at different positions on the retina of each eye. The

differences of the position between these two projections is the binocular disparity. This can be the strongest depth cue for close objects.

3.2.3 binocular depth cues: Binocular depth cues stem from binocular disparity and the state of the oculomotor system.

3.2.4 cardboard effect: The cardboard effect occurs when the imaging and display conditions reduce the reproduction magnification ratio of depth directions and distort the perception of objects with visually imperceptible thickness. The 3-D positions of stereoscopic objects are perceived stereoscopically but they appear unnaturally thin.

3.2.5 depth cues: Visual indicators of the relative distances between the observer and viewed objects. There are binocular and monocular depth cues.

3.2.6 depth distortion: Depth distortion is the discrepancy between the displayed depth reconstruction and real world. The cardboard effect is one example of a depth distortion.

3.2.7 depth motion: Objects moving in depth (e.g., toward the viewer).

3.2.8 depth resolution: spatial resolution in depth direction.

3.2.9 frame effect: 3D pictures appear highly unnatural when objects positioned in front of the screen approach the screen frame. This unnatural effect is called "the frame effect". The effect is generally reduced with a larger screen, because observers are less conscious of the existence of the frame when the screen is larger.

3.2.10 medical signs: Medical signs are physiological reactions that are measured and interpreted within a controlled procedure.

3.2.11 medical symptoms: Medical symptoms are expressed by the observer and constitute an indication of his or her perceived mental or physical state.

3.2.12 planar motion: Motion with constant depth (i.e., within a plane perpendicular to the observer).

3.2.13 puppet theatre effect: This is an example of size distortion. The reproduction magnification ratio of an object at the shooting distance (the perceived size) varies with the imaging and display conditions. The resulting distortion in size may make an object be perceived as unnaturally small.

3.2.14 size distortion: Distortions in the 3D geometry may occur due to inconsistencies between the capturing situation (notably the camera settings) and the reproduction (notably the display settings). The puppet theatre effect is an example.

3.2.15 monocular depth cues: Monocular depth cues can be perceived from a single eye's view. They include occlusion (objects hidden behind other objects), relative sizes of known objects, vanishing point perception, blur from the focus, motion parallax, light, shade, texture gradient, aerial perspective, height in the visual field, kinetic depth and others.

3.2.16 visual comfort zone: Visual comfort zone is defined as the depth interval which allows for 3D viewing without the introduction of visual discomfort.

3.2.17 visual discomfort: Visual discomfort is a negative sensation that is subjectively reported. Visual discomfort may be triggered by the perception of 3D presentations. Visual discomfort is usually a transient state that disappears quickly after the person stops watching the problematic stimuli. The amount and duration of visual discomfort is highly individual. Visual discomfort may have other causes, such as visual flicker.

3.2.18 visual fatigue: Visual fatigue is generally caused by the repetition of excessive visual effort and can build up as these efforts are repeated. It disappears after sufficient rest. Visual fatigue is a medical state that may be objectively diagnosed by the presence of objective indications

(medical signs) and subjective indications (medical symptoms). Visual fatigue can only be diagnosed by a medical doctor.

3.2.19 window violation: Window violation occurs when an object that is in front of the display overlaps the border of the screen. The 3D screen is often perceived as a window and in the real world objects can only be cropped when they lie behind the window. However, objects with crossed disparity are also cropped. When this occurs, the part of the object within the screen appears to be in front of the screen, while the cropping conflicts with this perception, by implying that the object should be at the depth of the screen or behind the screen.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

3DTV	Three Dimensional Television
ABV	Abnormal Binocular Vision
HD	High Definition
HRC	Hypothetical Reference Circuit
HUD	Head-up Display
PC	Pair Comparison
SD	Standard Definition

5 Conventions

"3D" in this Recommendation refers to a technology that projects dedicated views for each eye. The ultimate goal is that these views depict the same situation as would be seen in reality (e.g., horizontally shifted). The technology as a whole can include more than two views, as per a multi-view autostereoscopic display. This can be noted as S3D for stereoscopic TV, or 3DMV for multiview.

6 Technological issues for 3D displaying

This clause describes background information required to understand technological issues for displaying 3D content, including constraints of camera capture, transmission, display and the human visual system.

6.1 Accommodation-vergence conflict

Accommodation (eyes focussing on an object) and convergence (eyes pointing toward the object) are normally linked. Current 3D display technologies cause a conflict because accommodation and convergence do not follow the usual pattern. With 3D displays, the eyes focus on the screen but the eyes converge to a point in front of or behind the screen. This is the accommodation-vergence conflict.

The accommodation-vergence conflict has a static component (as described above) and also has a temporal component due to a link within the vision system. The accommodation control system and the vergence ocular motor system are linked. Changes in accommodation automatically induce changes in vergence and vice versa. This is the temporal component of the accommodation-vergence conflict. This is in conflict with current 3D display technology, because the accommodation always stays on the screen while the vergence changes according to the virtual object's position (binocular disparity).

More research is needed, because [b-Shiomi] indicates that there is no accommodation-vergence conflict.

6.2 3D geometric distortions

A disagreement between monocular and binocular depth cues may lead to visual discomfort. Disparity is the most important binocular depth cue.

In case of multi-view transmission, including stereoscopic transmissions, the resolution of the image impacts on the precision of the horizontal disparity and thus of the depth resolution. Likewise, in multi-view plus depth representations, the resolution of the depth map limits the picture quality of synthesized views.

The quality of the depth rendering is affected by both recording and display conditions which are closely linked.

In order to respect the 3D geometry of the scene during the reconstitution, it is necessary to have an agreement between the different depth cues (monocular and binocular) involved in the scenes. For example in the case of a cardboard effect, the monocular depth cues describe a continuous variation of depth but the binocular depth cues show less variation of depth than the monocular ones. These contradictions between the different depth cues are something unnatural and the scene looks layered.

To ensure respect of 3D geometry, it is then necessary to consider the characteristics of the video camera and the target display at the moment of the capture to ensure that there are no conflicts between monocular and binocular depth cues [b-Chen2] [b-Woods]. The distortion of the depth rendering can be evaluated as follows:

$$D_s = \frac{vb}{Bz + Mfb \times \left(1 - \frac{z}{d_{cov}}\right)}$$

With:

- f : the focal length of the camera
- b : the inter-camera baseline (distance between focal length)
- w_{ccd}, h_{ccd} : camera sensor width and height
- d_{cov} : the convergence distance
- D_s : depth distortion
- (x, y, z) : the position of a point in camera space (in front of the camera)
- B : inter-pupil baseline: the distance between the eyes of the observers
- W_{screen} : the width of the display
- M : the magnification factor: $M = W_{screen} / w_{ccd}$

These factors should be taken into account during capture. If not, conflict between depth cues will be revealed and distortion of the 3D geometry of objects will appear.

Figure 1 and Figure 2 provide examples of different capture and display settings with their rendering on a specific display [b-Woods].

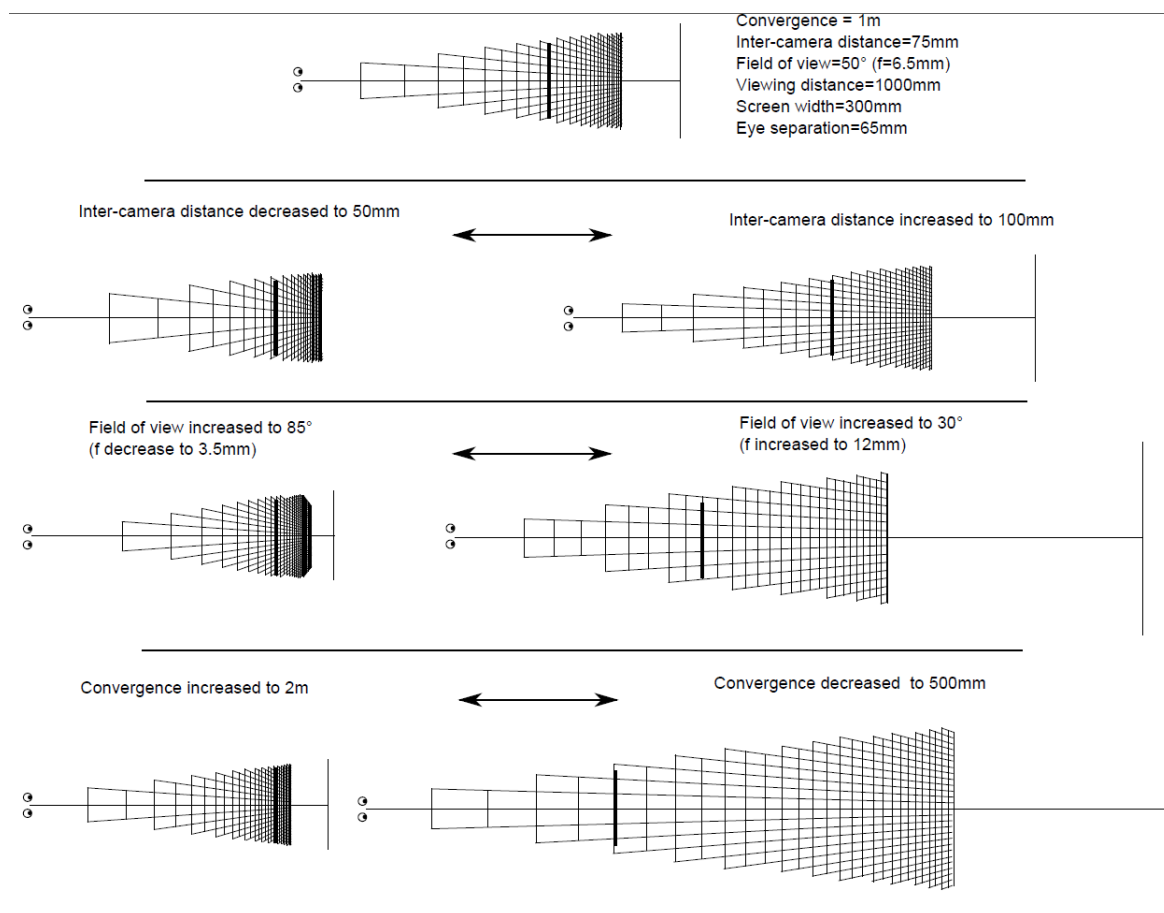


Figure 1 – Examples of different capture settings and their respective effects on the depth rendering on a fixed display [b-Woods]