

Рекомендация МСЭ-R ВТ.500-15

(05/2023)

Серия ВТ: Радиовещательная служба
(телевизионная)

**Методики субъективной оценки качества
телевизионных изображений**



Предисловие

Роль Сектора радиосвязи заключается в обеспечении рационального, справедливого, эффективного и экономичного использования радиочастотного спектра всеми службами радиосвязи, включая спутниковые службы, и проведении в неограниченном частотном диапазоне исследований, на основании которых принимаются Рекомендации.

Всемирные и региональные конференции радиосвязи и ассамблеи радиосвязи при поддержке исследовательских комиссий выполняют регламентарную и политическую функции Сектора радиосвязи.

Политика в области прав интеллектуальной собственности (ПИС)

Политика МСЭ-R в области ПИС излагается в общей патентной политике МСЭ-T/МСЭ-R/ИСО/МЭК, упоминаемой в Резолюции МСЭ-R 1. Формы, которые владельцам патентов следует использовать для представления патентных заявлений и деклараций о лицензировании, представлены по адресу: <http://www.itu.int/ITU-R/go/patents/ru>, где также содержатся Руководящие принципы по выполнению общей патентной политики МСЭ-T/МСЭ-R/ИСО/МЭК и база данных патентной информации МСЭ-R.

Серии Рекомендаций МСЭ-R

(Представлены также в онлайн-форме по адресу: <http://www.itu.int/publ/R-REC/ru>.)

Серия	Название
BO	Спутниковое радиовещание
BR	Запись для производства, архивирования и воспроизведения; пленки для телевидения
BS	Радиовещательная служба (звуковая)
BT	Радиовещательная служба (телевизионная)
F	Фиксированная служба
M	Подвижные службы, служба радиоопределения, любительская служба и относящиеся к ним спутниковые службы
P	Распространение радиоволн
RA	Радиоастрономия
RS	Системы дистанционного зондирования
S	Фиксированная спутниковая служба
SA	Космические применения и метеорология
SF	Совместное использование частот и координация между системами фиксированной спутниковой службы и фиксированной службы
SM	Управление использованием спектра
SNG	Спутниковый сбор новостей
TF	Передача сигналов времени и эталонных частот
V	Словарь и связанные с ним вопросы

Примечание. – Настоящая Рекомендация МСЭ-R утверждена на английском языке в соответствии с процедурой, изложенной в Резолюции МСЭ-R 1.

Электронная публикация
Женева, 2024 г.

© ITU 2024

Все права сохранены. Ни одна из частей данной публикации не может быть воспроизведена с помощью каких бы то ни было средств без предварительного письменного разрешения МСЭ.

РЕКОМЕНДАЦИЯ МСЭ-R ВТ.500-15

Методики субъективной оценки качества телевизионных изображений¹

(Вопрос МСЭ-R 102-4/6)

(1974-1978-1982-1986-1990-1992-1994-1995-1998-1998-2000-2002-2009-2012-2019-2023)

Сфера применения

В настоящей Рекомендации содержится описание методик оценки качества изображения, включая общие методы испытаний; шкалы, используемые при оценке качества; а также условия просмотра, рекомендуемые для проведения оценок. Рекомендация состоит из трех частей.

- Часть 1 – описание общих требований к проведению субъективной оценки качества телевизионных изображений и руководство по выбору конкретных методик, предназначенных для использования в различных ситуациях.
- Часть 2 – описание рекомендуемых методик оценки, которые могут применяться для субъективной оценки качества изображения.
- Часть 3 – описание методик, соответствующих конкретным форматам изображения и применениям согласно спецификациям, приведенным в частях 1 и 2.

Ключевые слова

Субъективная оценка, оценка изображения

Ассамблея радиосвязи МСЭ,

учитывая,

- a)* что накоплен большой объем информации о методах оценки качества изображения, используемых в различных лабораториях;
- b)* что изучение этих методов показывает существенный уровень согласия между различными лабораториями по ряду аспектов методик субъективной оценки;
- c)* что для обмена информацией между различными лабораториями важно принять стандартизированные методики оценки;
- d)* что некоторые методики, рекомендованные для лабораторных оценок, могут быть также использованы инженерами службы контроля при обычной или эксплуатационной оценке качества изображения и/или ухудшений с использованием пятибалльной шкалы оценки качества и ухудшений, производимой в обычных условиях эксплуатации или во время специальных работ;
- e)* что ввиду постоянного появления новых телевизионных сигналов, методов обработки сигналов и новых или усовершенствованных телевизионных служб могут потребоваться различные методики субъективной оценки изображения;
- f)* что введение таких методов обработки, сигналов и служб увеличит вероятность того, что качество на каждом участке пути следования сигнала в большей степени будет зависеть от процессов, происходящих в предыдущих частях этого пути,

рекомендует,

- 1 что для оценки качества изображения при лабораторных экспериментах и, если возможно, при практической работе следует использовать основные методики испытаний, шкалы оценок и условия просмотра, описанные в части 1;

¹ Настоящую Рекомендацию следует довести до сведения 12-й Исследовательской комиссии МСЭ-Т.

2 что, несмотря на существование альтернативных методик и разработку новых методик, когда это уместно, должны использоваться методы, описанные в части 2;

3 что в лабораторных экспериментах и, по возможности, в практической работе следует использовать общие методики испытаний, шкалы оценок и условия просмотра для оценки качества изображения конкретной системы изображений или конкретного применения, описанные в части 3;

4 что для упрощения обмена информацией между различными лабораториями следует соблюдать требования выбранной методики испытания, как описано в части 2;

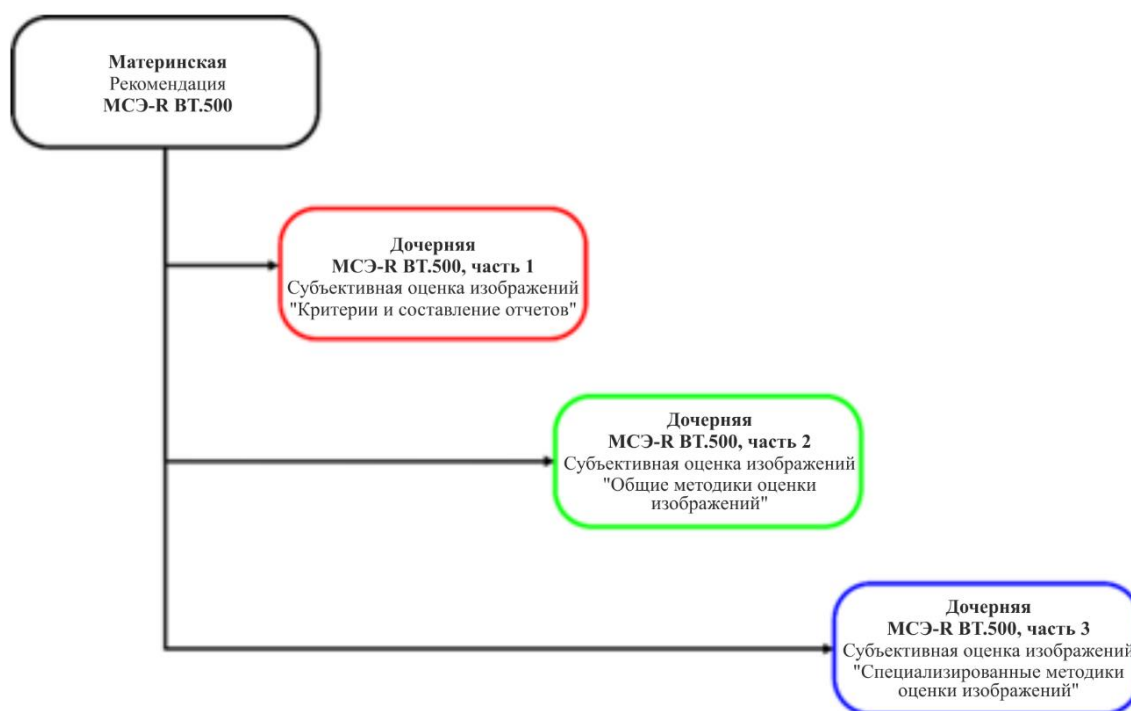
5 что для упрощения обмена информацией между различными лабораториями собранная информация должна быть статистически обработана по методике, подробно описанной в Приложении 2 к части 1;

6 что ввиду важности создания основы для субъективных оценок изображения, во всех отчетах о результатах испытаний должно быть представлено по возможности полное описание конфигурации испытаний, материалов для испытания, наблюдателей и методов.

Примечания по структуре и использованию настоящей Рекомендации (информационные)

Рекомендация МСЭ-R ВТ.500 состоит из трех полуавтономных частей в составе настоящей материнской Рекомендации, как показано на рисунке 1.

РИСУНОК 1
Структура Рекомендации МСЭ-R ВТ.500



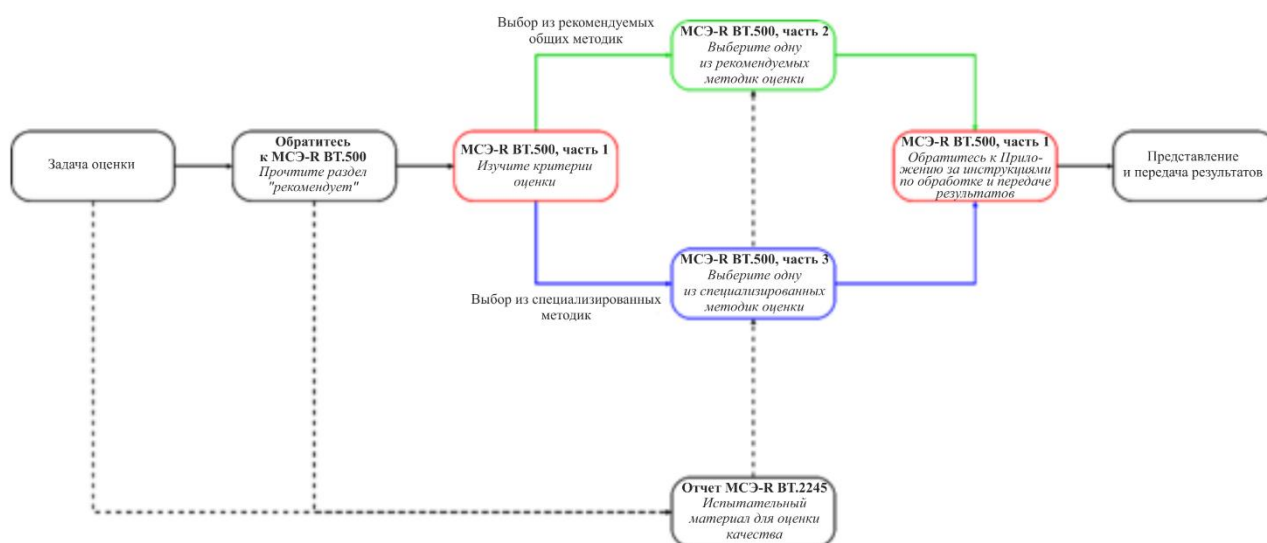
Лабораториям, желающим провести субъективную оценку изображений, для того чтобы выбрать наиболее подходящую методику процедур оценки, рекомендуется ознакомиться с вышеизложенным разделом *рекомендует* и использовать критерии, подробно описанные в части 1. В части 2 представлен обзор нескольких рекомендуемых методик субъективной оценки изображений, которые возможно использовать. В части 3 представлены сведения о некоторых дополнительных специализированных методиках, которые могут помочь при подготовке процедур субъективной оценки изображений в некоторых применениях.

Совет по использованию Рекомендации МСЭ-R ВТ.500

На рисунке 2 показан возможный рабочий процесс с использованием Рекомендации МСЭ-R ВТ.500.

РИСУНОК 2

Использование Рекомендации МСЭ-R ВТ.500



ВТ.0500-02

Обоснование

Структура частей этой версии Рекомендации МСЭ-R ВТ.500 позволяет добавлять новые и пересматривать существующие методики субъективной оценки изображений без необходимости выпуска новых Рекомендаций, чтобы не повторять одну и ту же информацию в нескольких документах, или выпуска пересмотров частей, изменения которых не требуется.

Другие Рекомендации по оценке изображения

Следующие Рекомендации относятся к объективному измерению качества изображения; в них могут предлагаться другие специализированные методики оценки изображения с использованием некоторых критериев оценки из МСЭ-R ВТ.500.

- | | |
|----------------------------|--|
| Рекомендация МСЭ-R ВТ.1683 | Объективные методы измерения воспринимаемого качества для цифрового вещательного телевидения стандартной четкости при наличии эталонного сигнала с полной полосой частот |
| Рекомендация МСЭ-R ВТ.1866 | Методы объективного измерения воспринимаемого качества изображения для радиовещательных применений с использованием телевидения низкой четкости при наличии полного эталонного сигнала |
| Рекомендация МСЭ-R ВТ.1867 | Технологии управления объективным качеством визуального восприятия для радиовещательных приложений, использующих стандарт ТВ малой четкости в присутствии эталонной полосы с уменьшенной шириной |

Рекомендация МСЭ-R ВТ.1885	Методы объективного измерения воспринимаемого качества изображения, предназначенные для цифрового вещательного телевидения стандартной четкости, при наличии эталонного сигнала с уменьшенной полосой частот
Рекомендация МСЭ-R ВТ.1907	Методы объективного измерения воспринимаемого качества изображения для радиовещательных применений с использованием ТВЧ при наличии полного эталонного сигнала
Рекомендация МСЭ-R ВТ.1908	Методы объективного измерения качества изображения для радиовещательных применений с использованием ТВЧ при наличии эталонного сигнала с ухудшенными характеристиками

ЧАСТЬ 1

Обзор требований к субъективной оценке изображения

1 Введение

Методы субъективной оценки изображения применяются для определения качества телевизионных систем с использованием таких измерений, которые непосредственно выражают реакцию тех, кто мог бы смотреть программу, полученную при помощи испытываемой системы. С этой точки зрения, понятно, что невозможно полностью охарактеризовать качество системы путем объективных измерений, следовательно, последние необходимо дополнять субъективными измерениями.

Как правило, существуют два класса субъективных оценок. Первый – это оценки, которые определяют качество системы в оптимальных условиях работы; такие оценки обычно называют оценками качества. Второй – это оценки, которые определяют способность системы поддерживать качество работы в неоптимальных условиях работы, обусловленных условиями передачи или распространения сигнала; такие оценки обычно называют оценками ухудшений.

Для проведения наиболее подходящих субъективных оценок в первую очередь необходимо выбрать из разных имеющихся вариантов методики, наилучшим образом соответствующую конкретным условиям и целям требуемой оценки изображения.

Для того чтобы облегчить этот выбор, следует рассмотреть общие характеристики, подробно описанные в разделе 2. Это позволит определить наиболее подходящие варианты, имеющие отношение к оцениваемой задаче или процессу.

После изучения этих вариантов следует обратиться к разделу 3 части 1, где представлен обзор рекомендуемых методик оценки изображения, которые можно использовать для выбора наиболее подходящей методики для оцениваемой задачи или процесса с учетом типа привлекаемых оценщиков и условий оценки.

Выбор наиболее подходящей методики, тем не менее, зависит от задач обслуживания испытываемой системы. Поэтому полное описание процедур оценки для конкретных применений приводится в части 2 и в других Рекомендациях МСЭ-R.

2 Общие особенности оценки

Здесь представлены общие условия просмотра для проведения субъективных оценок. Конкретные условия просмотра в целях проведения субъективных оценок конкретных систем представлены в соответствующих методиках.

ПРИМЕЧАНИЕ. – При субъективной оценке изображений с большим динамическим диапазоном рекомендуется обратиться к другим документам, ссылки на которые приведены в соответствующем разделе².

2.1 Общие условия просмотра

Лабораторная среда просмотра предназначена для обеспечения критических условий проверки систем. Общие условия просмотра для проведения субъективных оценок в лабораторной среде представлены в пункте 2.1.1.

Домашняя среда просмотра предназначена для предоставления средств в целях оценки качества на стороне ТВ цепочки потребителя. Общие условия просмотра, изложенные в пункте 2.1.2, воспроизводят домашнюю среду. Эти параметры были выбраны для установления условий, слегка более критических, чем типовые ситуации просмотра в домашней среде.

² В дальнейшем, по мере накопления опыта работы с изображениями с большим динамическим диапазоном, настоящая Рекомендация будет пересматриваться в целях включения дополнительных указаний.

2.1.1 Общие условия просмотра для субъективных оценок в лабораторной среде

Условия просмотра для оценщиков должны быть следующими:

- | | | |
|----|--|--|
| a) | Освещенность помещения: | низкая |
| b) | Цветность фона: | D_{65} |
| c) | Пиковая яркость ³ : | 70–250 кд/м ² (см. пункт 2.1.6.5) |
| d) | Коэффициент контрастности дисплея: | $\leq 0,02$ (см. пункт 2.1.6.4) |
| e) | Отношение яркости фона за дисплеем, на который выводится изображение, к пиковой яркости изображения: | $\approx 0,15$ |

2.1.2 Общие условия просмотра для субъективных оценок в домашней среде

- | | | |
|----|--|--|
| a) | Внешняя освещенность на экране (свет от внешних источников, падающий на экран, измеряется перпендикулярно к экрану): | 200 люкс |
| b) | Пиковая яркость: | 70–500 кд/м ² (см. пункт 2.1.6.4) |
| c) | Отношение яркости неактивного экрана к пиковой яркости дисплея: | $\leq 0,02$ (см. пункт 2.1.6.4) |

2.1.3 Расстояние просмотра

Расстояние просмотра выбирается в зависимости от размеров экрана согласно двум критериям: предпочтительное расстояние просмотра (PVD) и расчетное расстояние просмотра (DVD). Выбор одного из этих двух критериев будет зависеть от цели исследования.

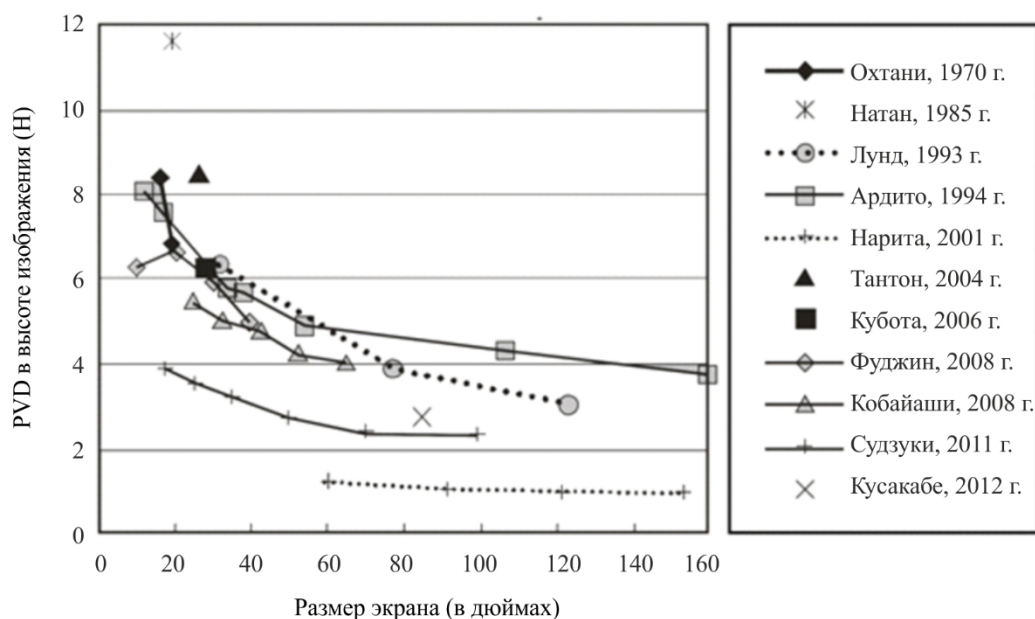
2.1.3.1 Предпочтительное расстояние просмотра

Предпочтительное расстояние просмотра (PVD) зависит от предпочтений зрителя, которые определяются эмпирическим путем. PVD (в зависимости от размеров экрана) показано на рисунке 1-1, который содержит ряд наборов данных, полученных из имеющихся источников. Эту информацию можно использовать при проектировании испытаний для субъективной оценки.

³ Пиковая яркость должна быть скорректирована в соответствии с освещенностью помещения.

РИСУНОК 1-1

Предпочтительное расстояние просмотра в зависимости от размеров экрана



ВТ.0500-01-1

2.1.3.2 Расчетное расстояние просмотра

Расчетное расстояние просмотра (DVD), или оптимальное расстояние просмотра, для цифровых систем – это расстояние, на котором два соседних пикселя составляют в глазу зрителя угол, равный 1 угловой минуте; а оптимальный горизонтальный угол просмотра – это угол, при котором изображение видно на оптимальном для него расстоянии просмотра.

В таблице 1-1 приведены оптимальные расстояния просмотра (и оптимальные горизонтальные углы просмотра) для нескольких систем разрешения изображения, выраженные в кратных высотах изображения.

ТАБЛИЦА 1-1

Оптимальный горизонтальный угол просмотра, оптимальное расстояние просмотра в высотах изображения (H)

Система изображения	Справочный документ	Формат экрана	Соотношение размеров элементов изображения	Оптимальный горизонтальный угол просмотра	Оптимальное расстояние просмотра
720 × 483	МСЭ-R ВТ.601	4:3	0,89	11°	7 H
640 × 480	VGA	4:3	1	11°	7 H
720 × 576	МСЭ-R ВТ.601	4:3	1,07	13°	6 H
1 024 × 768	XGA	4:3	1	17°	4,5 H
1 280 × 720	МСЭ-R ВТ.1543 и ВТ.1874	16:9	1	21°	4,8 H
1 400 × 1 050	SXGA+	4:3	1	23°	3,3 H
1 920 × 1 080	МСЭ-R ВТ.709	16:9	1	31°	3,2 H
3 840 × 2 160	МСЭ-R ВТ.2020	16:9	1	58°	1,6 H
7 680 × 4 320	МСЭ-R ВТ.2020	16:9	1	96°	0,8 H

Примечание. – При оценке изображения с учетом оптического разрешения для форматов 7680 × 4320 и 3840 × 2160 следует использовать меньшее значение расстояния просмотра. Когда оптическое разрешение не оценивается, можно использовать любое расстояние просмотра в диапазоне (для формата 3840 × 2160: от 1,6 до 3,2 высоты изображения; для формата 7680 × 4320: от 0,8 до 3,2 высоты изображения).

2.1.4 Угол обзора

Максимальный угол обзора по отношению к перпендикуляру должен быть ограничен таким образом, чтобы отклонения в воспроизведенном цвете на экране не были визуально заметны для наблюдателя. Оптимальный горизонтальный угол просмотра испытуемой системы изображения должен учитываться при определении угла обзора. Подробнее см. в пункте 1.8 Отчета МСЭ-R ВТ.2129.

2.1.5 Условия помещения – цветовая гамма

Цвет фона дисплея должен совпадать с эталонной белой точкой; для остальных поверхностей помещения следует использовать темные матовые поверхности. Цель состоит в том, чтобы минимизировать яркость рассеянного света на экране дисплея.

2.1.6 Дисплей

Использование дисплеев с разными характеристиками приведет к разному качеству субъективной оценки изображения. Поэтому настоятельно рекомендуется заранее проверить характеристики используемых дисплеев. Когда для субъективной оценки используются профессиональные плоскоэкранные дисплеи, можно обращаться к Рекомендации МСЭ-R ВТ.1886 "Эталонная функция электронно-оптического преобразования для плоскоэкранных дисплеев, используемых в студийном производстве программ ТВЧ" и Отчету МСЭ-R ВТ.2129 "Требования пользователей к плоскоэкранному дисплею (FPD) как основному дисплею в условиях производства программ ТВЧ".

В Отчете МСЭ-R ВТ.2390 содержится информация о лабораторных и бытовых дисплеях и условиях просмотра для оценки изображений с большим динамическим диапазоном (HDR).

2.1.6.1 Обработка дисплея

Обработка дисплея – например, масштабирование изображения, преобразование частоты кадров, усиление яркости изображения, – если реализована, должна производиться таким образом, чтобы избежать искажения изображения. Обработка HDR должна соответствовать HDR-системе, которая оценивается или используется во время оценки. При оценке условий восприятия или распространения это может означать использование соответствующих статических или динамических метаданных. В примечания к оценкам следует включать полную информацию о таких метаданных, чтобы другие лаборатории могли в точности повторить эти процедуры оценки.

При использовании для субъективных оценок изображения бытовых дисплеев важно, чтобы были отключены все параметры обработки изображения, если только предметом оценки не является воздействие такой обработки изображения.

При оценке изображений с чересстрочной разверткой в отчете о результатах испытаний следует указать, использовался ли преобразователь чересстрочной развертки в прогрессивную. Желательно не использовать этот преобразователь, если сигналы с чересстрочной разверткой могут быть отображены без него.

2.1.6.2 Разрешение дисплея

Разрешение профессиональных дисплеев обычно соответствует требуемым стандартам субъективной оценки в пределах их рабочего диапазона яркости.

Может быть предложено проверять и подтверждать максимальное и минимальное значения разрешения (центр и углы экрана) при используемом значении яркости.

Если для проведения субъективных оценок используются бытовые плоскоэкранные ТВ-дисплеи, то настоятельно рекомендуется проверять и подтверждать максимальное и минимальное значения разрешения (центр и углы экрана) при используемом значении яркости.

В настоящее время наиболее практичной системой, имеющейся в распоряжении тех, кто выполняет субъективные оценки в целях проверки разрешения дисплеев или бытовых телевизоров, является использование разворачиваемой испытательной таблицы, которая создается электронным способом.

2.1.6.3 Настройка дисплея

Яркость и контрастность дисплея должны быть настроены с учетом освещенности окружающей среды с использованием сигнала PLUGE в соответствии с Рекомендацией МСЭ-R ВТ.814.

Для оценки изображений со стандартным динамическим диапазоном (SDR) коэффициент контрастности дисплея следует измерять в соответствии с Рекомендацией МСЭ-R ВТ.815. При оценке изображений HDR следует обращаться к Отчету МСЭ-R ВТ.2390.

2.1.6.4 Контрастность дисплея

Существенное влияние на контрастность может оказывать освещенность окружающей среды.

В профессиональных дисплеях редко используются технологии повышения их контрастности в условиях сильной освещенности, поэтому, возможно, они не соответствуют требуемому стандарту контрастности при использовании в условиях сильной освещенности.

В бытовых дисплеях обычно используются технологии для получения большей контрастности в условиях сильной освещенности.

2.1.6.5 Яркость дисплея

При настройке яркости жидкокристаллических дисплеев желательно использовать регулировку интенсивности фоновой подсветки, а не масштабирование уровней сигнала с целью сохранения битовой точности. В случае других технологий дисплея, не использующих фоновую подсветку, уровень "белого" сигнала должен быть скорректирован другими способами, за исключением масштабирования уровней сигнала. Следует отметить, что плазменная панель регулирует яркость количеством световых излучений, и если установлена более низкая яркость, то воспроизведение цветовых тонов ухудшится.

2.1.6.6 Искажения движения на дисплее

Дисплей не должен вводить искажения движения в результате использования конкретных технологий дисплеев. С другой стороны, эффекты движения, включенные во входной сигнал, должны быть отображены на дисплее. При использовании бытовых дисплеев важно, чтобы ВСЕ функции обработки движущегося изображения были отключены.

2.1.6.7 Безопасные области изображения широкоэкранных дисплеев формата 16:9

Безопасные области изображения дисплеев формата 16:9 указаны в Рекомендации МСЭ-R ВТ.1848.

2.2 Исходные сигналы

Исходный сигнал представляет собой эталонное изображение и является входным сигналом для испытываемой системы. Его качество должно быть оптимальным для используемого телевизионного стандарта. Отсутствие дефектов в эталонной части представляемой пары сигналов является решающим фактором для получения стабильных результатов.

Неподвижные изображения и видеопоследовательности, сохраняемые в цифровой форме, могут быть воспроизведены наибольшее число раз, и, следовательно, они являются предпочтительными источниками сигналов. Лаборатории могут обмениваться такими сигналами для того, чтобы сделать сравнение систем более значимым.

Нередко требуется учесть, каким образом параметры испытываемой системы могли бы быть изменены в ходе обработки, выполненной на более ранних стадиях существования сигнала. Следовательно, желательно, чтобы испытание проводилось по возможности на каждом участке пути сигнала, где вероятно появление искажений, обусловленных обработкой, хотя бы и незаметных. Полученный в результате сигнал должен быть прозрачно записан и затем доступен для последующих испытаний при нисходящем потоке, если желательно проверить, каким образом искажения, возникающие из-за каскадной обработки, могут накапливаться во время прохождения сигнала по цепи. Такие записи должны храниться в библиотеке материала для испытаний для использования, при необходимости, в будущем, и вместе с ними должно храниться подробное описание истории записанного сигнала. При

необходимости источниками неподвижных изображений могут служить 35-мм сканеры слайдов. Достижимое разрешение соответствует требуемому для оценки качества обычного телевидения. Колориметрия и другие характеристики пленки могут привести к появлению субъективных различий в изображениях, полученных со студийной камеры. Если это оказывает влияние на результаты, то следует использовать непосредственно студийные сигналы, несмотря на то что это часто менее удобно. Как правило, для получения максимально хорошего субъективного качества изображения сканеры слайдов должны иметь возможность перенастройки при смене изображения.

Оценки характеристик обработки нисходящего цифрового потока часто выполняются с использованием цветовой маски. В студийных условиях такая маска очень чувствительна к освещенности студии. Следовательно, при оценках желательно использовать специальную пару слайдов с цветовой маской, которые дадут более точные результаты. Если необходимо, на передний план слайда может быть введено движение.

2.3 Выбор материалов для испытания

Существует множество подходов к тому, какие типы материала для испытания требуются для оценок качества телевидения. Однако на практике для каждого типа задач оценки должны быть использованы специальные виды материалов для испытания. Обзор типичных задач оценки и материалов для испытания, используемых для их решения, приведен в таблице 1-2.

ТАБЛИЦА 1-2

Выбор материалов для испытания*

Задачи оценки	Используемый материал
Общее качество со средним материалом	Обычный, "критичный, но не чрезмерно"
Пропускная способность, критичные применения (например, подача программ, постобработка и т. д.)	Набор, включающий очень критичный материал для испытываемых применений
Качество "адаптивных" систем	Материал, очень критичный для используемых "адаптивных" схем
Определение слабостей и возможных улучшений	Материал, критичный к определенным атрибутам
Определение причин, по которым системы выглядят различными	Широкий набор разнообразных материалов
Преобразование между различными стандартами	Критичный к различиям (например, скорости полей)

* Понятно, что все возможные материалы для испытаний могут быть частью содержания телевизионной программы. Более полное руководство по выбору материалов для испытаний содержится в Приложениях 3 и 4.

Некоторые параметры могут дать увеличение обычного порядка искажений для большинства изображений или последовательностей. В таких случаях результаты, полученные при очень небольшом количестве изображений или последовательностей (например, двух), могут предоставить достаточную оценку.

Однако новые системы часто подвержены влиянию, которое сильно зависит от содержания изображения или последовательности. В таких случаях будет существовать статистическое распределение вероятности искажений и содержания изображений или испытательных последовательностей для всего времени демонстрации программ. Без знания формы этого распределения, что обычно имеет место, следует очень осторожно производить отбор материала для испытаний и интерпретация результатов.

Как правило, очень важно включать критичный материал, так как его можно учесть при интерпретации результатов, однако невозможно осуществлять экстраполяцию исходя из некритичных материалов. В тех случаях, когда содержание сцены или последовательности изображений влияет на результаты, для испытываемой системы должен быть выбран материал "критичный, но не чрезмерно". Фраза "не чрезмерно" означает, что изображение может быть, предположительно, одной из частей ТВ программы, демонстрируемой в обычные часы. В таких случаях следует использовать, по крайней мере, четыре объекта, из которых, например, половина очень критична и половина умеренно критична.

2.3.1 Испытательные последовательности МСЭ-R

Многие организации разработали испытательные неподвижные изображения и последовательности. Отчет МСЭ-R ВТ.2245 "Испытательные материалы для оценки качества изображения ТВЧ и ТСВЧ, включая HDR-TV" содержит подробное описание материалов для испытания ТВЧ и ТСВЧ, которые можно использовать для субъективной оценки. Дополнительные идеи по выбору испытательных материалов приведены в Приложениях 1 и 2 к части 1 настоящей Рекомендации.

2.4 Диапазон условий и привязки

Поскольку большинство методов оценки чувствительны к изменениям диапазона и распределения условий наблюдения, сеансы экспертизы должны включать в себя все возможные диапазоны факторов. Однако они могут быть приближены к более ограниченному диапазону путем представления также условий, приходящихся на крайние точки шкалы. Эти условия могут быть представлены как отдельные примеры и определены как самые крайние (прямая привязка) или распределены по всему сеансу и не определены как наиболее крайние (косвенная привязка).

2.5 Наблюдатели

Наблюдатели могут быть или не быть экспертами, что определяется задачами оценки. Наблюдатель-эксперт – это наблюдатель, имеющий опыт определения искажений изображения, которые может внести испытываемая система. Не являющийся экспертом ("необученный") наблюдатель – это наблюдатель, не имеющий опыта определения искажений изображения, которые может внести испытываемая система. В любом случае наблюдатели не должны – в настоящее время или ранее – быть непосредственными участниками разработки испытываемой системы, то есть в такой степени, чтобы приобрести специальные и подробные знания.

2.5.1 Число наблюдателей

Если в выбранной методике не указано иное, следует привлекать к участию не менее 15 наблюдателей. Необходимое число наблюдателей зависит от критичности и надежности принятой процедуры испытания и от ожидаемого воздействия оцениваемого эффекта. Для исследований с ограниченной сферой охвата, например экспериментального характера, можно привлекать менее 15 наблюдателей. В этом случае исследования должны быть обозначены как "неформальные". Следует указать уровень компетенции наблюдателей в области оценки качества телевизионного изображения.

2.5.2 Отбор наблюдателей

Как правило, до сеанса следует оценить (или скорректировать) остроту зрения наблюдателей при помощи таблицы Снеллена или Ландольта и проверить цветовое зрение, используя специально отобранные таблицы (например, Ишихара).

В разделах А1-2.3 и А1-2.4 подробно описаны различные сценарии отбора наблюдателей, которые возможно применять для различных методик испытаний. Если лабораторные или менее формальные испытания проводятся в рамках программы испытаний в нескольких местах или организациях, важно, чтобы полная информация о методе и критериях отбора наблюдателей была распространена и включена в публикуемые результаты.

Как правило, следует как можно более подробно указывать характеристики своих групп оценки, эти данные могут включать профессиональную категорию (например, работник радиовещательной организации, студент университета, служащий), пол и возрастную группу.

ПРИМЕЧАНИЕ. – Исследование согласованности результатов различных испытательных лабораторий показало, что между результатами, полученными в различных испытательных лабораториях, могут появляться систематические различия. Такие различия будут иметь особенное значение, если предлагается объединить результаты, предоставленные несколькими различными лабораториями с целью повышения чувствительности и надежности эксперимента.

Вероятное объяснение различий результатов разных лабораторий состоит в возможном различии в уровнях навыков разных групп оценщиков. Необходимо провести дополнительное исследование для оценки обоснованности этой гипотезы, и, если она окажется верной, количественно оценить изменения, обусловленные данным фактором.

2.5.3 Инструкции для оценки

Оценщикам следует внимательно ознакомиться с методикой оценки, типами искажений или качественными показателями, которые могут появиться, шкалой оценок, последовательностью и порядком сеанса. С этой целью используются тренировочные последовательности, демонстрирующие диапазон и тип искажений, которые должны быть оценены, причем демонстрационные изображения в этих последовательностях должны отличаться от тех, что будут предложены в испытаниях, но производить сравнимое с ними впечатление. В случае оценок качества оно может быть определено как совпадение определенных воспринимаемых атрибутов.

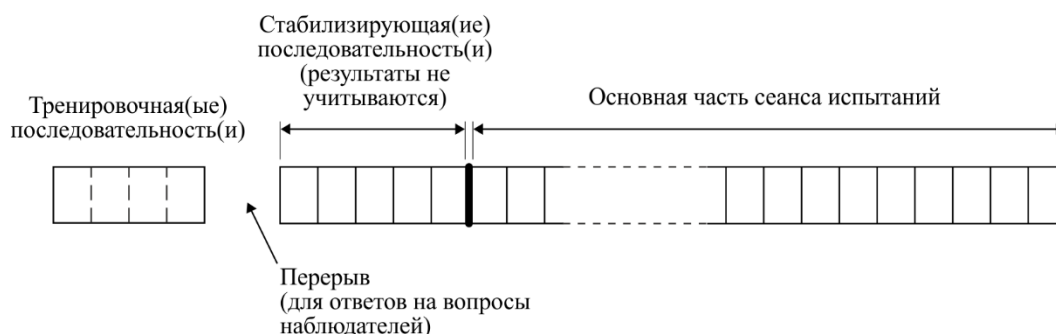
2.6 Сеанс испытаний

Продолжительность сеанса не должна превышать получаса. В начале первого сеанса следует провести примерно пять "тренировочных демонстраций" для стабилизации мнения наблюдателей. Информация, полученная в ходе этих демонстраций, не должна учитываться в результатах испытания. Если необходимо проведение нескольких сеансов, то в начале каждого следующего сеанса необходимо проведение только трех "тренировочных демонстраций".

Для демонстраций должен быть использован случайный порядок (например, полученный из греко-латинских квадратов); но порядок условий проведения испытаний должен быть построен так, чтобы любые воздействия усталости или привыкания на оценку были сбалансированы от сеанса к сеансу. Некоторые демонстрации могут быть повторены в разных сеансах для проверки согласованности.

РИСУНОК 1-2

Структура представления сеанса испытания



ВТ.0500-01-2

2.7 Представление результатов

Поскольку результаты изменяются в широких пределах, решения, принятые в большинстве методов оценки, нецелесообразно толковать в абсолютных значениях (например, качество изображения или последовательности изображений).

Для каждого испытательного параметра должны быть приведены среднее значение и 95% доверительный интервал статистического распределения классов оценок. Если оценивалось изменение искажения в зависимости от изменения значения параметра, то следует использовать методы аппроксимации кривых. Логистическая аппроксимация кривых и логарифмическая ось позволят получить представление в виде прямой линии. Это предпочтительная форма представления результатов. В Приложении 1 к части 1 настоящей Рекомендации приведена более подробная информация об обработке данных.

Результаты должны быть представлены совместно со следующей информацией:

- подробное описание конфигурации испытания;
- подробное описание материалов для испытания;
- тип источника изображения и отображающих дисплеев (см. Примечание 1);
- количество и тип оценщиков (см. Примечание 2);
- используемые эталонные системы;

- средняя оценка эксперимента;
- исходные и поправленные величины средней оценки и 95% доверительный интервал в том случае, если мнение одного или нескольких наблюдателей не принималось в расчет в соответствии с процедурой, описанной ниже.

ПРИМЕЧАНИЕ 1. – Поскольку существуют некоторые данные о том, что размер устройства отображения может оказывать влияние на результаты субъективных оценок, экспериментаторам предлагается подробно описывать размер экрана, марку и номер модели используемых устройств отображения при проведении любых экспериментов.

ПРИМЕЧАНИЕ 2. – Существуют данные о том, что различия в уровне навыков групп, осуществляющих просмотр (даже среди групп, в состав которых не входят эксперты), может оказывать влияние на результаты субъективных оценок просмотров. Для содействия проведению дальнейшего исследования этого фактора экспериментаторам предлагается как можно более подробно указывать характеристики своих групп просмотра. Соответствующие показатели могут включать: возрастной и гендерный состав участников группы или образование или профессиональную категорию участников группы.

3 Выбор методов испытаний

В телевизионных испытаниях используются самые разнообразные базовые методы испытаний. Однако на практике для решения конкретных задач оценки должны быть использованы определенные методы. В части 3 настоящей Рекомендации содержится руководство по субъективной оценке качества изображения в соответствующих форматах изображения и применениях.

Приложение 1 к части 1

Анализ и представление результатов

A1-1 Введение

В ходе субъективного эксперимента по оценке качества телевизионной системы осуществляется сбор большого объема данных. Эти данные в форме бланков оценок наблюдателей или их электронных эквивалентов должны быть преобразованы с помощью статистических методов для получения результатов в форме графиков и/или числового выражения/формулы/алгоритма, которые обобщают данные о качестве испытываемых систем.

Приведенный ниже анализ применим к результатам, полученным с использованием методов SS, метода DSIS и метода DSCQS для оценки качества телевизионного изображения, которые описаны в Приложениях 1, 2 и 3 к части 2 настоящей Рекомендации, и к другим альтернативным методам, в которых применяются цифровые шкалы. В первом и втором случаях ухудшение оценивается по пятибалльной или многобалльной шкале. В последнем случае используются непрерывные шкалы оценок, а результаты (различия оценок эталонного изображения и испытываемого реального изображения) приводятся к целочисленным значениям от 0 до 100.

A1-2 Общие методы анализа

В результате проведения испытаний в соответствии с принципами методов, которые описаны в разделе 2 части 1, будут получены распределения целочисленных значений, например от 0 до 5 или от 0 до 100. Эти распределения будут различаться из-за разницы в оценках наблюдателей и влияния разных условий, связанных с экспериментом, например, использования нескольких изображений или последовательностей.

Испытание будет состоять из ряда демонстраций *L*. Каждая демонстрация будет демонстрацией ряда условий испытаний *J*, применяемых к демонстрации ряда испытательных последовательностей/испытательных изображений *K*. В некоторых случаях каждое сочетание испытательной последовательности / испытательного изображения и условия испытаний может быть повторено несколько раз (*R*).

A1-2.1 Вычисление средних оценок

Первый этап анализа результатов – это вычисление средней оценки \bar{u}_{jkr} для каждой демонстрации:

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr}, \quad (1)$$

где:

u_{ijkr} : оценка наблюдателя i для условия испытания j , последовательности/изображения k , количества повторов r ;

N : число наблюдателей.

Аналогично могут быть вычислены общие средние оценки \bar{u}_j и \bar{u}_k для каждого условия испытания и каждой испытательной последовательности/испытательного изображения.

A1-2.2 Расчет доверительного интервала

A1-2.2.1 Обработка первоначальных (нескорректированных и/или неаппроксимированных) данных

При представлении результатов испытания все средние оценки должны иметь соответствующий доверительный интервал, который получают исходя из стандартного отклонения и размера каждой выборки.

Предлагается использовать 95%-ный доверительный интервал, который задается выражением:

$$\left[\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} + \delta_{jkr} \right], \quad (2)$$

где:

$$\delta_{jkr} = 1,96 \frac{S_{jkr}}{\sqrt{N}}. \quad (3)$$

Стандартное отклонение S_{jkr} для каждой демонстрации задается выражением:

$$S_{jkr} = \sqrt{\frac{\sum_{i=1}^N (\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}}. \quad (4)$$

С вероятностью 95% абсолютное значение разницы между экспериментальной средней оценкой и "верной" средней оценкой (при очень большом числе наблюдателей) меньше 95%-ного доверительного интервала при условии, что распределение отдельных оценок соответствует определенным требованиям.

Аналогично может быть вычислено стандартное отклонение S_j . Однако отмечается, что в случаях, когда используется меньшее количество испытательных последовательностей/испытательных изображений, на это стандартное отклонение будут больше влиять различия между используемыми испытательными последовательностями, чем различия между оценщиками, участвующими в оценке.

A1-2.2.2 Обработка скорректированных и/или аппроксимированных данных

Для данных, в отношении которых были скорректированы эффекты остаточного снижения/повышения качества и влияния краев шкалы оценок, или данных, представленных в форме характеристики искажений или закона добавления искажений после аппроксимации (ввиду зависимости экспериментальных средних оценок качества от этих искажений), доверительный интервал должен быть вычислен с использованием преобразований статистических переменных, учитывающих дисперсию соответствующей переменной.

Если результаты оценки качества представлены как характеристика искажений (т. е. экспериментальная кривая), то нижний и верхний доверительные пределы доверительного интервала будут функцией каждого экспериментального значения. Для вычисления этих доверительных пределов следует рассчитать стандартное отклонение и оценить аппроксимацию его зависимости для каждого экспериментального значения исходной характеристики ухудшений.

A1-2.3 Последующее отсеивание наблюдателей

A1-2.3.1 Последующее отсеивание на основе эксцесса для методов DSIS, DSCQS и альтернативных методов за исключением метода SSCQE

Прежде всего, необходимо выяснить, является ли распределение оценок в случае испытательной демонстрации нормальным или не является таковым, используя испытание β_2 (путем вычисления коэффициента эксцесса функции, то есть отношения момента четвертого порядка к квадрату момента второго порядка). Если β_2 лежит в пределах от 2 до 4, то распределение может считаться нормальным. По каждой демонстрации оценки u_{ijkr} каждого наблюдателя следует сравнить с соответствующим средним значением \bar{u}_{jkr} плюс соответствующее стандартное отклонение S_{jkr} , умноженное на два (если нормальное) или умноженное на $\sqrt{20}$ (если не нормальное) (P_{jkr}), и с соответствующим средним значением минус то же стандартное отклонение, умноженное на два или на $\sqrt{20}$ (Q_{jkr}). Каждый раз, когда оценка наблюдателя находится выше P_{jkr} , показание счетчика, связанного с каждым наблюдателем (P_i), увеличивается. Аналогично, каждый раз, когда оценка наблюдателя находится ниже Q_{jkr} , показание счетчика, связанного с каждым наблюдателем (Q_i), увеличивается. Наконец, должны быть вычислены следующие два отношения: $P_i + Q_i$, поделенное на общее количество оценок каждого наблюдателя за весь сеанс, и $P_i - Q_i$, поделенное на $P_i + Q_i$, как абсолютное значение. Если первое отношение больше 5%, а второе отношение меньше 30%, то наблюдатель i должен быть исключен (см. Примечание 1).

ПРИМЕЧАНИЕ. – Настоящая процедура не должна применяться более одного раза к результатам данного эксперимента. Кроме того, использование этой процедуры должно быть ограничено в случаях, когда используется относительно небольшое число наблюдателей (например, менее 20), и все они не являются экспертами.

Настоящая процедура рекомендуется при применении метода ЕСР (DSIS); она также с успехом применяется при использовании метода DSCQS и альтернативных методов.

Описанный выше процесс может быть выражен математически следующим образом:

По каждой испытательной демонстрации вычисляют среднее значение \bar{u}_{jkr} , стандартное отклонение S_{jkr} и коэффициент эксцесса β_{2jkr} , где β_{2jkr} задается выражением:

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \quad \text{с} \quad m_x = \frac{\sum_{i=1}^N (u_{ijkr} - \bar{u}_{ijkr})^x}{N}. \quad (5)$$

Для каждого наблюдателя i находят P_i и Q_i , то есть:

для $j, k, r = 1, 1, 1$ до J, K, R ,

если $2 \leq \beta_{2jkr} \leq 4$, то:

если $u_{ijkr} \geq \bar{u}_{jkr} + 2 S_{jkr}$, тогда $P_i = P_i + 1$;

если $u_{ijkr} \leq \bar{u}_{jkr} - 2 S_{jkr}$, тогда $Q_i = Q_i + 1$;

или же:

если $u_{ijkr} \geq \bar{u}_{jkr} + \sqrt{20} S_{jkr}$, тогда $P_i = P_i + 1$;

если $u_{ijkr} \leq \bar{u}_{jkr} - \sqrt{20} S_{jkr}$, тогда $Q_i = Q_i + 1$.

Если $\frac{P_i + Q_i}{J \cdot K \cdot R} > 0,05$ и $\left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0,3$, то исключается наблюдатель i ,

при этом:

N : число наблюдателей;

J : количество условий испытаний, включая эталонное условие;

K : количество испытательных изображений или последовательностей;

R: количество повторов;

L: количество испытательных демонстраций (в большинстве случаев количество демонстраций будет равно $J \cdot K \cdot R$, однако отмечается, что некоторые оценки могут проводиться при неодинаковом количестве последовательностей для каждого условия испытания).

A1-2.3.2 Последующее отсеивание на основе эксцесса для метода SSCQE

В случае отбора определенных наблюдателей при использовании процедуры испытания SSCQE область применения больше не является одной из схем испытания (сочетание условия испытания и испытательной последовательности), а является временным интервалом (например, 10-секундным сегментом вынесения решения) схемы испытания. Применяется двухступенчатая фильтрация – на первом шаге выявляются и исключаются наблюдатели, решения которых имеют сильное отклонение от среднего поведения, на втором шаге осуществляется выявление и отбор несоответствующих наблюдателей без какого-либо рассмотрения систематического отклонения.

Шаг 1. Обнаружение местных отклонений от нормального порядка

В этом случае также необходимо выяснить, является ли распределение оценок для каждого временного интервала каждой схемы испытания "нормальным" или "не нормальным", используя испытание β_2 . Если β_2 лежит между 2 и 4, распределение может считаться "нормальным". Затем процесс применяется к каждому временному интервалу каждой схемы испытания как математически выражено ниже.

Для каждого временного интервала каждой схемы испытания вычисляются среднее значение \bar{u}_{jkr} , стандартное отклонение S_{jklr} и коэффициент β_{2jklr} с использованием оценок u_{ijk} каждого наблюдателя. Коэффициент β_{2jklr} задается формулой:

$$\beta_{2jklr} = \frac{m_4}{(m_2)^2} \quad \text{с} \quad m_x = \frac{\sum_{n=1}^N (u_{njklr} - \bar{u})^x}{N}. \quad (6)$$

Для каждого наблюдателя i находят P_i и Q_i , то есть:

для $j, k, l, r = 1, 1, 1$ до J, K, L, R

если $2 \leq \beta_{2jklr} \leq 4$, то:

если $u_{njklr} \geq \bar{u}_{jklr} + 2 S_{jklr}$, тогда $P_i = P_i + 1$;

если $u_{njklr} \leq \bar{u}_{jklr} - 2 S_{jklr}$, тогда $Q_i = Q_i + 1$;

или же:

если $u_{njklr} \geq \bar{u}_{jklr} + \sqrt{20} S_{jklr}$, тогда $P_i = P_i + 1$;

если $u_{njklr} \leq \bar{u}_{jklr} - \sqrt{20} S_{jklr}$, тогда $Q_i = Q_i + 1$.

Если $\frac{P_i}{J \cdot K \cdot L \cdot R} > X\%$ или $\frac{Q_i}{J \cdot K \cdot L \cdot R} > X\%$, тогда исключается наблюдатель i ,

при этом:

N: число наблюдателей;

J: количество временных интервалов в рамках испытательной комбинации условия испытания и испытательной последовательности;

K: количество условий испытаний;

L: количество последовательностей;

R: количество повторов.

Данный процесс позволяет исключить наблюдателей, решения которых существенно отличаются от средних оценок. На рисунке 1-3 представлены два примера (две крайние кривые, демонстрирующие значительные отклонения). Однако данные критерии исключения не позволяют обнаружить возможные отклонения от нормального порядка, которые являются другим важным источником необъективности.

Шаг 2. Обнаружение местных отклонений оценок от нормального порядка

В случае шага 2 обнаружение также основано на формуле отбора, представленной в Приложении 2. Вносится небольшое изменение, касающееся области применения. Набор входных данных снова состоит из оценок, полученных по всем временным интервалам (например, 10 секунд) всех схем испытания. Однако на этот раз оценки предварительно концентрируются вокруг общего значения для сведения к минимуму эффекта отклонения, который уже был учтен на первом этапе процесса. Затем применяется обычный процесс.

Прежде всего, следует выяснить, является ли это распределение оценок для каждого временного интервала каждой схемы испытания "нормальным" или "не нормальным", используя испытание β_2 . Если β_2 лежит между 2 и 4, распределение может считаться "нормальным". Затем процесс применяется к каждому временному интервалу каждой схемы испытания, как математически выражено ниже.

Первый шаг процесса – это вычисление концентрированных оценок по каждому временному интервалу и каждому наблюдателю. Средняя оценка \bar{u}_{klr} для каждой схемы испытания определяется как:

$$\bar{u}_{klr} = \frac{1}{N} \cdot \frac{1}{J} \sum_{n=1}^N \sum_{j=1}^J u_{njklr} \quad (7)$$

Аналогично определяется средняя оценка для каждой схемы испытания и каждого наблюдателя:

$$\bar{u}_{nklr} = \frac{1}{J} \sum_{j=1}^J u_{njklr}, \quad (8)$$

а u_{njklr} соответствует оценке наблюдателя i для временного интервала j , условия испытания k , последовательности l , повтора r .

Для каждого наблюдателя концентрированные оценки u^*_{njklr} рассчитываются следующим образом:

$$u^*_{njklr} = u_{njklr} - \bar{u}_{nklr} + \bar{u}_{klr}. \quad (9)$$

Для каждого временного интервала каждой схемы измерения вычисляются среднее значение \bar{u}^*_{jklr} , стандартное отклонение S^*_{jklr} и коэффициент $\beta_2^*_{jklr}$, который задается формулой:

$$\beta_2^*_{jklr} = \frac{m_4}{(m_2)^2} \quad \text{при} \quad m_x = \frac{\sum_{n=1}^N (u^*_{njklr})^x}{N}. \quad (10)$$

Для каждого наблюдателя i находят P^*_i и Q^*_i , то есть:

для $j, k, l, r = 1, 1, 1, 1$ до J, K, L, R ,

если $2 \leq \beta_2^*_{jklr} \leq 4$, то:

если $u^*_{njklr} \geq \bar{u}^*_{jklr} + 2 S^*_{jklr}$, тогда $P^*_i = P^*_i + 1$;

если $u^*_{njklr} \leq \bar{u}^*_{jklr} - 2 S^*_{jklr}$, тогда $Q^*_i = Q^*_i + 1$;

или же:

если $u^*_{njklr} \geq \bar{u}^*_{jklr} + \sqrt{20} S^*_{jklr}$, тогда $P^*_i = P^*_i + 1$;

если $u^*_{njklr} \leq \bar{u}^*_{jklr} - \sqrt{20} S^*_{jklr}$, тогда $Q^*_i = Q^*_i + 1$.

Если $\frac{P_i^* + Q_i^*}{J \cdot K \cdot L \cdot R} > Y$ и $\left| \frac{P_i^* - Q_i^*}{P_i^* + Q_i^*} \right| < Z$, тогда исключается наблюдатель i ,

при этом:

- N : число наблюдателей;
- J : количество временных интервалов в рамках испытательной комбинации условия испытания и испытательной последовательности;
- K : количество условий испытаний;
- L : количество последовательностей;
- R : количество повторов.

Предлагаемыми значениями параметров (X , Y , Z), адаптированными к данному методу, по опыту, являются 0,2; 0,1; 0,3.

A1-2.3.3 Последующее отсеивание на основе корреляции

Каждый наблюдатель должен иметь стабильный и последовательный метод, обеспечивающий беспристрастную оценку ухудшения качества каждой сцены и алгоритма. Критерий отклонения позволяет подтвердить уровень последовательности оценок одного наблюдателя по сравнению со средней оценкой всех наблюдателей, принимавших участие в данной сессии испытаний. Критерий принятия решения основан на корреляции индивидуальных оценок с соответствующими средними оценками всех наблюдателей, принимавших участие в испытаниях. Эта процедура проще в реализации, чем соответствующий метод, описанный в предыдущих разделах.

A1-2.3.3.1 Корреляция Пирсона

Предполагается, что соотношение между шкалой качества и диапазоном оценок наблюдателей является линейным, поэтому можно применить корреляцию Пирсона.

Основная цель состоит в проверке простым методом того, согласуются ли оценки одного наблюдателя со средними оценками, выставленными всеми наблюдателями на протяжении всего сеанса испытания. Скрытый эталон считается привязкой высокого качества. Если используются слабая и сильная привязки, то они повышают оценку корреляции, с другой стороны, уменьшаются разбросы корреляции между наблюдателями.

$$r(x, y) = \frac{\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right) \left(\sum_{i=1}^n y_i\right)}{n}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}\right) \left(\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}\right)}}, \quad (11)$$

где:

- x : средняя оценка всех наблюдателей по трем элементам (алгоритм, скорость передачи данных в битах, сцена);
- y : индивидуальная оценка одного наблюдателя по тем же трем элементам;
- n : (число алгоритмов) \times (число сцен);
- i : {номер кодека, номер скорости передачи данных по битам, номер сцены}.

А1-2.3.3.2 Ранговая корреляция Спирмена

Ранговая корреляция Спирмена может применяться, даже если предполагается, что соотношение между шкалой качества и диапазоном оценок наблюдателей не является линейным⁴:

$$r(x, y) = \left[1 - \frac{6 \times \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n^3 - n} \right], \quad (12)$$

где:

- x_i : средняя оценка всех наблюдателей по трем элементам (алгоритм, скорость передачи данных в битах, цена);
- y_i : индивидуальная оценка одного наблюдателя по тем же трем элементам;
- n : (число алгоритмов) \times (число сцен);
- $R(x_i$ или $y_i)$: порядок ранжирования;
- i : {номер кодека, номер скорости передачи данных по битам, номер сцены}.

А1-2.3.3.3 Критерий обоснования окончательного отказа для исключения наблюдателя из испытания

Корреляции Пирсона и Спирмена используются для исключения наблюдателя(ей), в соответствии со следующими условиями:

ЕСЛИ $[\text{mean}(r) - \text{sdt}(r)] >$ порога максимальной корреляции (МСТ).

Порог отказа = порогу максимальной корреляции (МСТ).

В ИНОМ СЛУЧАЕ порог отказа = $[\text{mean}(r) - \text{sdt}(r)]$.

ЕСЛИ $[r$ (Наблюдатель)] $>$ порога отказа.

ТОГДА наблюдатель "i" не исключается из испытания.

В ИНОМ СЛУЧАЕ наблюдатель "i" исключается из испытания;

где:

- r = мин (корреляция Пирсона, ранговая корреляция Спирмена);
- $\text{mean}(r)$: среднее значение корреляций для всех участвующих в испытании наблюдателей;
- $\text{sdt}(r)$: стандартное отклонение корреляций для всех участвующих в испытании наблюдателей.

Порог максимальной корреляции (МСТ) = 0,85.

Значение МСТ = 0,85 действительно для методов SAMVIQ и DSCQS, в ином случае должно рассматриваться значение МСТ = 0,7 для методов SS и DSIS.

А1-2.4 Расчет средних баллов и доверительных интервалов в случае испытаний в сложных условиях

Очень часто субъективные испытания приходится проводить в сложных условиях. Например, в ходе краудсорсингового испытания испытуемые подвергаются воздействию менее контролируемой среды, чем в лаборатории. В крупномасштабных испытаниях, проводимых несколькими лабораториями, межлабораторная вариация может вызвать значительное расхождение полученных оценок. Методы, представленные в разделах А1 2.1–А1-2.3, зачастую не вполне подходят для таких условий. В настоящем разделе представлен усовершенствованный метод анализа данных, который, как было показано, улучшает качество восстановленных средних оценок и доверительных интервалов. В Прилагаемом документе 1 к настоящему Приложению содержится эталонная реализация на языке Python.

⁴ Обычно результаты корреляции Пирсона и результаты корреляции Спирмена очень близки.

Идея, лежащая в основе этого метода, заключается в следующем. Полезно явно моделировать поведение каждого участника; в частности, двумя важными человеческими факторами, влияющими на решения участников, являются необъективность и последовательность участников. С помощью итеративной процедуры этот метод определяет общую оценку истинного качества каждой демонстрации, а также необъективность и последовательность каждого участника. Возможная интерпретация оцениваемого истинного качества каждой демонстрации – "средняя экспертная оценка, взвешенная по последовательности, без необъективности". По сравнению с последующим отсеиванием участников, описанным в разделе А1-2.3.1, которое либо сохраняет, либо отклоняет все решения участника ("жесткое отклонение"), этот метод можно охарактеризовать как "мягкое отклонение". То есть для "выпадающего" участника, который голосует непоследовательно, решения будут иметь небольшой вес и, следовательно, малый вклад в общую MOS. Побочным результатом этого метода является оценка необъективности и последовательности каждого участника. Это ценная информация о пригодности участника для проведения субъективных испытаний, поэтому ее возможно использовать для отбора участников для будущих испытаний. Например, если участник принимал крайне непоследовательные решения, он/она может быть исключен/исключена из будущих сессий.

Сначала определяются средние оценки для каждой демонстрации по всем участникам и повторам:

$$\bar{u}_{jk} = \frac{1}{N \cdot R} \sum_{i=1}^N \sum_{r=1}^R u_{ijkr}, \quad (13)$$

где u_{ijkr} – оценка наблюдателя i для условия j , последовательности/изображения k , повтора r , N – число наблюдателей, R – количество повторов.

На втором этапе оценивается необъективность каждого наблюдателя b_i следующим образом:

$$b_i = \frac{1}{J \cdot K \cdot R} \sum_{j=1}^J \sum_{k=1}^K \sum_{r=1}^R u_{ijkr} - \bar{u}_{jk}, \quad (14)$$

где J и K – количество условий и количество последовательностей, соответственно. Затем выполняются следующие шаги в итерационном цикле.

Текущая оценка средней оценки по каждой демонстрации записывается как \bar{u}_{jk}^c , то есть:

$$\bar{u}_{jk}^c = \bar{u}_{jk} \quad (15)$$

с последующим вычислением остатка в каждой полученной оценке, который не может быть объяснен средней оценкой и необъективностью наблюдателя:

$$e_{ijkr} = u_{ijkr} - \bar{u}_{jk} - b_i. \quad (16)$$

Эти остатки далее используются для расчета непоследовательности каждого наблюдателя σ_i следующим образом:

$$\sigma_i = \sqrt{\frac{1}{J \cdot K \cdot R} \sum_{j=1}^J \sum_{k=1}^K \sum_{r=1}^R (u_{ijkr} - \mu_{e_i})^2}, \quad (17)$$

где:

$$\mu_{e_i} = \frac{1}{J \cdot K \cdot R} \sum_{j=1}^J \sum_{k=1}^K \sum_{r=1}^R e_{ijkr}. \quad (18)$$

Новые оценки средних оценок далее возможно получить следующим образом:

$$\bar{u}_{jk} = \frac{\sum_{i=1}^N \sum_{r=1}^R \sigma_i^{-2} (u_{ijkr} - b_i)}{\sum_{i=1}^N \sum_{r=1}^R \sigma_i^{-2}}, \quad (19)$$

далее оценки необъективности обновляются согласно уравнению (12).

Цикл завершается, если:

$$\sum_{j=1}^J \sum_{k=1}^K (\bar{u}_{jk} - \bar{u}_{jk}^c)^2. \quad (20)$$

После завершения определяется стандартное отклонение оценки для каждой демонстрации следующим образом:

$$S_{jk} = \frac{\sigma_j}{\sqrt{N}}, \quad (21)$$

где:

$$\sigma_j = \sqrt{\frac{1}{N \cdot R} \sum_{i=1}^N \sum_{r=1}^R (e_{ijk_r} - \mu_{e_j})^2} \quad (22)$$

и

$$\mu_{e_j} = \frac{1}{N \cdot R} \sum_{i=1}^N \sum_{r=1}^R e_{ijk_r}. \quad (23)$$

Далее рассчитывается окончательный доверительный интервал в соответствии с уравнениями (2) и (3).

A1-3 Обработка для нахождения зависимости между средней оценкой и объективным измерением искажений изображения

Если субъективные испытания проводились с целью изучения взаимосвязи между объективным измерением искажения и средними оценками \bar{u} (\bar{u} вычислено в соответствии с пунктом A1-2.1), то может быть полезным следующий процесс, заключающийся в нахождении простой непрерывной зависимости между \bar{u} и параметром искажений.

A1-3.1 Аппроксимация при помощи симметричной логической функции

Особый интерес представляет аппроксимация этой экспериментальной зависимости логической функцией.

Обработка данных \bar{u} может быть осуществлена следующим образом.

Шкала значений \bar{u} нормирована путем использования непрерывной переменной p , так что

$$p = (\bar{u} - u_{min}) / (u_{max} - u_{min}), \quad (24)$$

где:

u_{min} : минимальная оценка, существующая на u -шкале для худшего качества;

u_{max} : максимальная оценка, существующая на u -шкале для лучшего качества.

Графическое представление зависимости между p и D показывает, что кривая стремится к форме сигмоида с симметричным наклоном, при условии, что естественные пределы значений D расположены достаточно далеко от области, в которой u быстро изменяется.

Функция $p = f(D)$ теперь может быть аппроксимирована разумно выбранной логической функцией, задаваемой общим соотношением:

$$p = 1 / [1 + \exp (D - D_M) G], \quad (25)$$

где D_M и G – постоянные величины, а G может быть положительной или отрицательной величиной.

Значение p , полученное исходя из оптимальной аппроксимации логической функции, используется для предоставления численного значения I , получаемого в соответствии с соотношением:

$$I = (1/p - 1). \quad (26)$$

Значения D_M и G могут быть получены исходя из экспериментальных данных после следующего преобразования:

$$I = \exp (D - D_M) G. \quad (27)$$

Это выражение устанавливает линейную связь путем использования логарифмической шкалы для I :

$$\log_e I = (D - D_M) G. \quad (28)$$

Интерполяция прямой линией является простой, а в некоторых случаях ее точности достаточно для того, чтобы считать ее представляющей искажение вследствие влияния, измеряемого D .

Тогда наклон характеристики выражается формулой:

$$S = \frac{D_M - D}{\log_e I} = \frac{1}{G}, \quad (29)$$

которая дает оптимальное значение G . D_M – это значение D для $I = 1$.

Прямая линия может быть названа характеристикой искажений, связанной с конкретным рассматриваемым искажением. Будет отмечено, что прямая линия может быть определена характеристическими значениями D_M и G логической функции.

A1-3.2 Аппроксимация несимметричной функцией

A1-3.2.1 Описание функции

Аппроксимация зависимости между экспериментальными оценками и объективной количественной оценкой искажения изображения с помощью симметричной логической функцией является наиболее результативной в случае, когда параметр искажения D может быть измерен в относительных единицах, например S/N (дБ). Если параметр искажения измеряется физической единицей d , например временной задержкой (мс), то соотношение (27) следует заменить формулой:

$$I = (d/d_M)^{1/G}, \quad (30)$$

и поэтому уравнение (25) принимает вид:

$$p = 1 / \left[1 + (d/d_M)^{1/G} \right]. \quad (31)$$

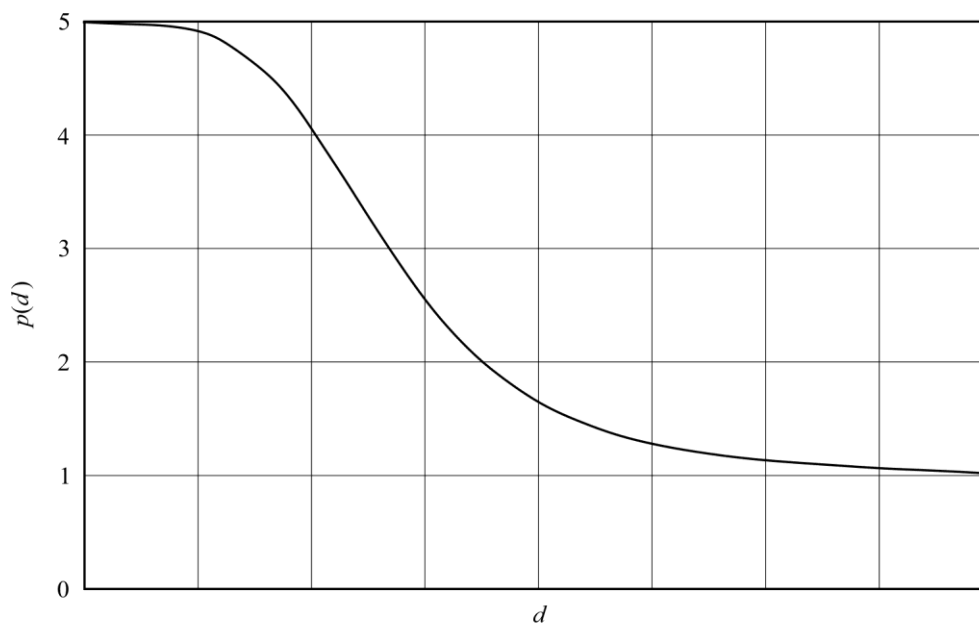
Данная функция аппроксимирует логическую функцию несимметричным способом.

A1-3.2.2 Оценка параметров аппроксимации

Оценка оптимальных параметров функции, которая обеспечивает минимальные остаточные ошибки между реальными данными и функцией, может быть получена при использовании какого-либо рекурсивного алгоритма оценок. На рисунке 1-3 показан пример использования несимметричной функции для представления реальных субъективных данных. Это представление позволяет получить оценку конкретных объективных количественных оценок, соответствующих интересующему субъективному значению: например, 4,5 по пятибалльной шкале.

РИСУНОК 1-3

Несимметричная аппроксимация



А1-3.3 Устранение остаточного ухудшения/улучшения и влияния краев шкалы

На практике использование логической функции иногда не позволяет избежать некоторых различий между экспериментальными данными и аппроксимацией. Эти расхождения могут быть обусловлены влияниями края шкалы или одновременным наличием нескольких искажений в испытании, что может оказать влияние на статистическую модель и исказить форму теоретической логической функции.

Была выявлена разновидность влияния края шкалы, заключающаяся в том, что наблюдатели стремятся не использовать крайние значения шкалы оценок, в частности в отношении оценок высокого качества. Это может происходить из-за ряда факторов, в том числе психологического нежелания делать крайние суждения. Кроме того, использование среднего арифметического значения оценок в соответствии с уравнением (1) вблизи краев шкалы может обусловить наличие необъективных результатов ввиду негауссовского распределения решений в этих областях.

Часто в испытаниях указывают остаточное ухудшение (даже в отношении эталонных изображений средняя оценка достигает только значения $\bar{u}_0 < u_{max}$).

Существует несколько полезных подходов для коррекции первоначальных данных оценок с целью обработки действительных выводов (см. таблицу 1-3).

Устранение влияний краев, если они существуют на экспериментальных данных, является частью обработки данных, которая имеет большое значение. Отметим, что в этих процедурах коррекции задействованы специальные предположения, поэтому рекомендуется применять их с осторожностью, а об их использовании следует сообщать при представлении результатов.

ТАБЛИЦА 1-3

Сравнение методов устранения влияний краев шкалы

Методы компенсации влияний краев	Свойства		
	Компенсация остаточного ухудшения качества	Компенсация остаточного улучшения качества	Сдвиг к центру шкалы
Без компенсации	Нет	Нет	Нет
Преобразование линейной шкалы	Да	Может быть существенная ошибка	Нет
Преобразование нелинейной шкалы ⁽¹⁾	Да	Да	Нет
Метод, основанный на добавлении вставок	Да	Нет	Да
Мультипликативный метод	Да	Нет	Да

⁽¹⁾ В соответствии с нелинейным преобразованием шкалы следует рассчитать скорректированные оценки:

$$u_{corr} = C(\bar{u} - u_{mid}) + u_{mid}$$

$$C = \frac{\bar{u} - u_{0min}}{u_{0max} - u_{0min}} \frac{u_{max} - u_{mid}}{u_{0max} - u_{mid}} + \frac{u_{0max} - \bar{u}}{u_{0max} - u_{0min}} \frac{u_{min} - u_{mid}}{u_{0min} - u_{mid}},$$

где:

- u_{corr} : скорректированная оценка;
- \bar{u} : нескорректированная экспериментальная оценка;
- u_{min} , u_{max} : края шкалы оценивания;
- u_{mid} : середина шкалы оценивания;
- u_{0min} , u_{0max} : нижний и верхний пределы в тенденции экспериментальных оценок.

А1-3.4 Включение аспекта надежности в графики

Из средних оценок для каждого испытанного искажения и соответствующих 95%-ных доверительных интервалов составляется три ряда оценок:

- ряд минимальных оценок (средние значения – доверительные интервалы);
- ряд средних оценок;
- ряд максимальных оценок (средние значения + доверительные интервалы).

Полученные по этим трем рядам параметры оценки затем оцениваются независимо. Три полученные функций затем могут быть отображены на том же графике: две – на основании максимального и минимального рядов – пунктирными линиями и одна – средняя оценка – сплошной линией. Экспериментальные значения тоже отмечаются на этом графике (см. рисунок 1-4). Следовательно, получается оценка 95%-ной непрерывной доверительной области.

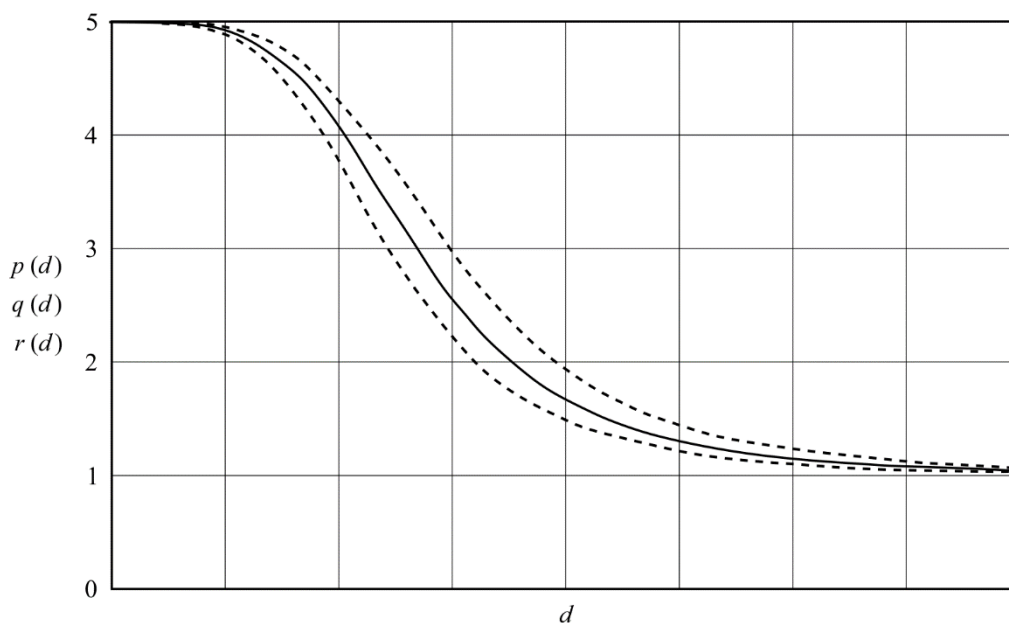
Таким образом, для оценки 4,5 (порог заметности для данного метода) мы можем непосредственно из графика найти 95%-ный доверительный интервал, который может быть использован с целью определения диапазона допусков.

Промежуток между кривыми для максимальных и минимальных оценок – это не 95%-ный интервал, а их средняя оценка.

Не менее 95% экспериментальных значений должны попадать в доверительный интервал, в ином случае может быть сделан вывод о том, что при проведении испытания возникла проблема или что выбранная модель функции не является оптимальной.

РИСУНОК 1-4

Случай несимметричной характеристики искажений



$p(d)$: ряд средних оценок

$q(d)$: ряд минимальных оценок

$r(d)$: ряд максимальных оценок

d : объективная количественная оценка искажений

A1-4 Выводы

Описана процедура оценки доверительных интервалов, то есть точности набора испытаний с использованием субъективных оценок.

Процедура позволяет также получить оценку средних общих величин, соответствующих не только конкретному рассматриваемому эксперименту, но и другим экспериментам, проводимым по той же методике.

Поэтому такие величины могут быть использованы для построения диаграмм поведения доверительных интервалов, полезных в отношении субъективных оценок, а также для планирования будущих экспериментов.

Прилагаемый документ 1 к Приложению 1

Эталонная реализация метода, описанного в разделе A1-2.4

Настоящий Прилагаемый документ включает написанную на языке Python эталонную реализацию метода анализа данных, представленного в разделе A1-2.4. Используемый код и примеры данных также общедоступны в пакете SUREAL Python по адресу: https://github.com/Netflix/sureal/tree/master/itur_bt500_demo.

Входные данные подготовлены следующим образом. Необработанные оценки организованы в двумерную матрицу и разделены запятыми. Каждая строка соответствует демонстрации (исходное изображение, просматриваемое при испытаниях), каждый столбец соответствует участнику.

Необязательно, чтобы каждый участник ставил оценку каждой демонстрации. Если участник i не поставил оценку демонстрации jk , на место (jk,i) вставляется "nan" (не число). Входные данные помещены в файл .csv. Ниже приведен пример небольшого файла .csv, который содержит оценки 20 участников по 30 демонстрациям, которые проигрывались два раза.

```
5.0,nan,5.0,4.0,2.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
1.0,3.0,5.0,2.0,5.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
3.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,3.0,4.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
1.0,4.0,3.0,4.0,5.0,5.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0
4.0,5.0,nan,3.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
4.0,3.0,2.0,5.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
1.0,3.0,4.0,5.0,1.0,4.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,5.0,3.0,5.0,5.0,4.0,3.0,5.0
3.0,5.0,4.0,2.0,4.0,5.0,4.0,5.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0
5.0,2.0,1.0,3.0,3.0,4.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
1.0,2.0,1.0,1.0,3.0,1.0,1.0,1.0,1.0,3.0,1.0,2.0,2.0,1.0,1.0,1.0,2.0,1.0,1.0,2.0
5.0,5.0,3.0,1.0,3.0,1.0,2.0,2.0,2.0,3.0,2.0,3.0,4.0,2.0,1.0,2.0,2.0,1.0,2.0,2.0
5.0,2.0,4.0,3.0,4.0,2.0,2.0,2.0,2.0,4.0,3.0,3.0,3.0,5.0,2.0,2.0,2.0,4.0,2.0,2.0
5.0,5.0,5.0,5.0,4.0,3.0,3.0,3.0,3.0,5.0,3.0,4.0,4.0,3.0,2.0,2.0,3.0,3.0,3.0,3.0
5.0,5.0,4.0,3.0,5.0,4.0,4.0,4.0,4.0,5.0,4.0,4.0,5.0,4.0,3.0,3.0,4.0,3.0,3.0,4.0
1.0,4.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,3.0
1.0,4.0,1.0,4.0,3.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,5.0,4.0
4.0,2.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0
2.0,5.0,3.0,2.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
5.0,5.0,5.0,5.0,3.0,3.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0
4.0,5.0,5.0,3.0,5.0,2.0,2.0,3.0,1.0,3.0,3.0,2.0,3.0,5.0,1.0,1.0,2.0,2.0,2.0,2.0
```

1.0,2.0,2.0,4.0,5.0,1.0,2.0,2.0,1.0,3.0,2.0,2.0,4.0,2.0,3.0,1.0,2.0,2.0,1.0,3.0
4.0,5.0,3.0,5.0,2.0,3.0,2.0,3.0,3.0,4.0,2.0,3.0,4.0,3.0,3.0,1.0,2.0,2.0,2.0,3.0
1.0,5.0,3.0,5.0,4.0,2.0,3.0,3.0,3.0,5.0,3.0,3.0,4.0,2.0,3.0,2.0,3.0,3.0,2.0,3.0
5.0,5.0,5.0,5.0,1.0,4.0,4.0,3.0,3.0,5.0,3.0,4.0,4.0,4.0,4.0,3.0,4.0,3.0,3.0,4.0
5.0,5.0,5.0,5.0,4.0,5.0,4.0,4.0,4.0,5.0,5.0,4.0,4.0,5.0,5.0,5.0,5.0,3.0,4.0,4.0
5.0,1.0,4.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0
3.0,4.0,4.0,2.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
4.0,1.0,3.0,5.0,3.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0
3.0,3.0,1.0,3.0,1.0,1.0,2.0,3.0,1.0,3.0,1.0,3.0,1.0,2.0,2.0,2.0,2.0,2.0,2.0,2.0
5.0,3.0,2.0,2.0,5.0,3.0,1.0,3.0,1.0,4.0,3.0,4.0,3.0,4.0,3.0,3.0,3.0,2.0,1.0,2.0
,

5.0,nan,5.0,4.0,2.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
1.0,3.0,5.0,2.0,5.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
3.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,3.0,4.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
1.0,4.0,3.0,4.0,5.0,5.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0
4.0,5.0,nan,3.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
4.0,3.0,2.0,5.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
1.0,3.0,4.0,5.0,1.0,4.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,5.0,3.0,5.0,5.0,4.0,3.0,5.0
3.0,5.0,4.0,2.0,4.0,5.0,4.0,5.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0
5.0,2.0,1.0,3.0,3.0,4.0,5.0,5.0,3.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0
1.0,2.0,1.0,1.0,3.0,1.0,1.0,1.0,1.0,3.0,1.0,2.0,2.0,1.0,1.0,1.0,2.0,1.0,1.0,2.0
5.0,5.0,3.0,1.0,3.0,1.0,2.0,2.0,2.0,3.0,2.0,3.0,4.0,2.0,1.0,2.0,2.0,1.0,2.0,2.0
5.0,2.0,4.0,3.0,4.0,2.0,2.0,2.0,2.0,4.0,3.0,3.0,3.0,5.0,2.0,2.0,2.0,4.0,2.0,2.0
5.0,5.0,5.0,5.0,4.0,3.0,3.0,3.0,3.0,5.0,3.0,4.0,4.0,3.0,2.0,2.0,3.0,3.0,3.0,3.0
5.0,5.0,4.0,3.0,5.0,4.0,4.0,4.0,4.0,5.0,4.0,4.0,5.0,4.0,3.0,3.0,4.0,3.0,3.0,4.0
1.0,4.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,4.0,5.0,4.0,5.0,5.0,3.0
1.0,4.0,1.0,4.0,3.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,4.0
4.0,2.0,5.0,5.0,4.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0
2.0,5.0,3.0,2.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
5.0,5.0,5.0,5.0,3.0,3.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0
4.0,5.0,5.0,3.0,5.0,2.0,2.0,3.0,1.0,3.0,3.0,2.0,3.0,5.0,1.0,1.0,2.0,2.0,2.0,2.0
1.0,2.0,2.0,4.0,5.0,1.0,2.0,2.0,1.0,3.0,2.0,2.0,4.0,2.0,3.0,1.0,2.0,2.0,1.0,3.0
4.0,5.0,3.0,5.0,2.0,3.0,2.0,3.0,3.0,4.0,2.0,3.0,4.0,3.0,3.0,1.0,2.0,2.0,2.0,3.0
1.0,5.0,3.0,5.0,4.0,2.0,3.0,3.0,3.0,5.0,3.0,3.0,4.0,2.0,3.0,2.0,3.0,3.0,2.0,3.0
5.0,5.0,5.0,5.0,1.0,4.0,4.0,3.0,3.0,5.0,3.0,4.0,4.0,4.0,4.0,3.0,4.0,3.0,3.0,4.0
5.0,5.0,5.0,5.0,4.0,5.0,4.0,4.0,4.0,5.0,5.0,4.0,4.0,5.0,5.0,5.0,5.0,3.0,4.0,4.0
5.0,1.0,4.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0
3.0,4.0,4.0,2.0,5.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,4.0,5.0,5.0,5.0,5.0,5.0,5.0,5.0
4.0,1.0,3.0,5.0,3.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0
3.0,3.0,1.0,3.0,1.0,1.0,2.0,3.0,1.0,3.0,1.0,3.0,1.0,2.0,2.0,2.0,2.0,2.0,2.0,2.0
5.0,3.0,2.0,2.0,5.0,3.0,1.0,3.0,1.0,4.0,3.0,4.0,3.0,4.0,3.0,3.0,3.0,2.0,1.0,2.0

Код на языке Python, реализующий этот метод, находится в файле *demo_bt500.py*.

demo_bt500.py:

```
import argparse
import csv
import sys
import pprint

import numpy as np
from scipy import linalg

def read_csv_into_3darray(csv_filepath):
    """
    Read data from CSV file.

    The data should be organized in a 2D matrix, separated by comma. Each row
    correspond to a PVS; each column corresponds to a subject. If a vote is
    missing, a 'nan' is put in place.

    If some subjects evaluated a PVS multiple times, another 2D matrix of the
    same size [num_PVS, num_subjects] can be added under the first one. A row
    with a single comma (,) should be placed before the repetition matrix.
    Where the repeated vote is not available, a 'nan' is put in place.

    :param csv_filepath: filepath to the CSV file.
    :return: the numpy array in 3D [num_PVS, num_subjects, num_repetitions].
    """

    data = []
    data3dlist = []
    with open(csv_filepath, 'rt') as datafile:
        datareader = csv.reader(datafile, delimiter=',')

        for row in datareader:
            if row != [",", ""]:
                data.append(np.array(row, dtype=np.float64))
            else:
                data3dlist.append(data)
                data = []
        data3dlist.append(data)

    data3d = np.zeros((len(data3dlist[0]), len(data3dlist[0][0]), len(data3dlist)))

    for r_idx, r_mat in enumerate(data3dlist):
        data3d[:, :, r_idx] = r_mat

    return data3d

def weighed_nanmean_2d(a, wts, axis):
    """
    Compute the weighted arithmetic mean along the specified axis, ignoring
    NaNs. It is similar to numpy's nanmean function, but with a weight.

    :param a: 1D array.
    :param wts: 1D array carrying the weights.
    :param axis: either 0 or 1, specifying the dimension along which the means
    are computed.
    :return: 1D array containing the mean values.
    """

    assert len(a.shape) == 2
    assert axis in [0, 1]
    d0, d1 = a.shape
    if axis == 0:
        return np.divide(
            np.nansum(np.multiply(a, np.tile(wts, (d1, 1)).T), axis=0),
            np.nansum(np.multiply(~np.isnan(a), np.tile(wts, (d1, 1)).T), axis=0)
```

```

    )
elif axis == 1:
    return np.divide(
        np.nansum(np.multiply(a, np.tile(wts, (d0, 1))), axis=1),
        np.nansum(np.multiply(~np.isnan(a), np.tile(wts, (d0, 1))), axis=1),
    )
else:
    assert False

def one_or_nan(x):
    """
    Construct a "mask" array with the same dimension as x, with element NaN
    where x has NaN at the same location; and element 1 otherwise.

    :param x: array_like
    :return: an array with the same dimension as x
    """
    y = np.ones(x.shape)
    y[np.isnan(x)] = float('nan')
    return y

def get_sos_j(sig_j, u_jkir):
    """
    Compute SOS (standard deviation of score) for presentation jk
    :param sig_j:
    :param u_jkir:
    :return: array containing the SOS for presentation jk
    """
    den = np.nansum(
        stack_3rd_dimension_along_axis(one_or_nan(u_jkir) / np.tile(sig_j ** 2,
        (u_jkir.shape[1], 1)).T[:, :, None],
        axis=1),
        axis=1)
    s_jk_std = 1.0 / np.sqrt(np.maximum(0., den))
    return s_jk_std

def stack_3rd_dimension_along_axis(u_jkir, axis):
    """
    Take the 3D input matrix, slice it along the 3rd axis and stack the resulting 2D
    matrices
    along the selected matrix while maintaining the correct order.
    :param u_jkir: 3D array of the shape [JK, I, R]
    :param axis: 0 or 1
    :return: 2D array containing the values
        - if axis=0, the new shape is [R*JK, I]
        - if axis = 1, the new shape is [JK, R*I]
    """
    assert len(u_jkir.shape) == 3
    JK, I, R = u_jkir.shape

    if axis == 0:
        u = np.zeros([R * JK, I])

        for r in range(R):
            u[r * JK:(r + 1) * JK, :] = u_jkir[:, :, r]

    elif axis == 1:
        u = np.zeros([JK, R * I])

        for r in range(R):
            u[:, r * I:(r + 1) * I] = u_jkir[:, :, r]

    else:
        NotImplementedError

    return u

```

```

def run_alternating_projection(u_jkir):
    """
    Run Alternating Projection (AP) algorithm.

    :param u_jkir: 3D numpy array containing raw votes. The first dimension
    corresponds to the presentation (jk); the second dimension corresponds to the
    subjects (i); the third dimension corresponds to the repetitions (r).
    If a vote is missing, the element is NaN.

    :return: dictionary containing results keyed by 'mos_j', 'sos_j', 'bias_i'
    and 'inconsistency_i'.
    """
    JK, I, R = u_jkir.shape

    # video by video, estimate MOS by averaging over subjects
    u_jk = np.nanmean(stack_3rd_dimension_along_axis(u_jkir, axis=1), axis=1) # mean
    marginalized over i

    # subject by subject, estimate subject bias by comparing with MOS
    b_jir = u_jkir - np.tile(u_jk, (I, 1)).T[:, :, None]
    b_i = np.nanmean(stack_3rd_dimension_along_axis(b_jir, axis=0), axis=0) # mean
    marginalized over j

    MAX_ITR = 1000
    DELTA_THR = 1e-8
    EPSILON = 1e-8

    itr = 0
    while True:

        u_jk_prev = u_jk

        # subject by subject, estimate subject inconsistency by averaging the
        # residue over stimuli
        e_jkir = u_jkir - np.tile(u_jk, (I, 1)).T[:, :, None] - np.tile(b_i, (JK, 1))[:,
        :, None]
        sig_i = np.nanstd(stack_3rd_dimension_along_axis(e_jkir, axis=0), axis=0)
        sig_j = np.nanstd(stack_3rd_dimension_along_axis(e_jkir, axis=1), axis=1)

        # video by video, estimate MOS by averaging over subjects, inversely
        # weighted by residue variance
        w_i = 1.0 / (sig_i ** 2 + EPSILON)
        # mean marginalized over i:
        u_jk = weighed_nanmean_2d(
            stack_3rd_dimension_along_axis(u_jkir - np.tile(b_i, (JK, 1))[:, :, None],
axis=1),
            wts=np.tile(w_i, R), # same weights for the repeated observations
            axis=1)

        # subject by subject, estimate subject bias by comparing with MOS,
        # inversely weighted by residue variance
        b_jir = u_jkir - np.tile(u_jk, (I, 1)).T[:, :, None]
        # mean marginalized over j:
        b_i = np.nanmean(stack_3rd_dimension_along_axis(b_jir, axis=0), axis=0)

        itr += 1

        delta_u_jk = linalg.norm(u_jk_prev - u_jk)

        msg = 'Iteration {itr:4d}: change {delta_u_jk}, u_jk {u_jk}, ' \
            'b_i {b_i}, sig_i {sig_i}'.format(
                itr=itr, delta_u_jk=delta_u_jk, u_jk=np.mean(u_jk),
                b_i=np.mean(b_i), sig_i=np.mean(sig_i))

        sys.stdout.write(msg + '\r')
        sys.stdout.flush()

```

```

    if delta_u_jk < DELTA_THR:
        break

    if itr >= MAX_ITR:
        break

    u_jk_std = get_sos_j(sig_j, u_jkir)
    sys.stdout.write("\n")

    mean_b_i = np.mean(b_i)
    b_i -= mean_b_i
    u_jk += mean_b_i

    return {
        'mos_j': list(u_jk),
        'sos_j': list(u_jk_std),
        'bias_i': list(b_i),
        'inconsistency_i': list(sig_i),
    }
}

if __name__ == "__main__":
    parser = argparse.ArgumentParser()

    parser.add_argument(
        "--input-csv", dest="input_csv", nargs=1, type=str,
        help="Filepath to input CSV file. The data should be organized in a 2D "
        "matrix, separated by comma. The rows correspond to PVSSs; the "
        "columns correspond to subjects. If a vote is missing, input 'nan'"
        " instead.", required=True)

    args = parser.parse_args()
    input_csv = args.input_csv[0]

    o_jir = read_csv_into_3darray(input_csv)

    ret = run_alternating_projection(o_jir)

    pprint.pprint(ret)

```

Приложение 2 к части 1

Описание общего формата взаимного обмена файлами данных

Назначение общего формата взаимного обмена файлами данных – содействие обмену данными между лабораториями, участвующими в совместной международной кампании по проведению субъективной оценки.

Любая экспертиза субъективной оценки проводится в соответствии с пятью последовательными и зависимыми этапами: подготовка к испытанию, проведение испытания, обработка данных, представление и интерпретация результатов. В случае широкомасштабных международных кампаний обычно бывает, что работа распределяется между различными участвующими лабораториями:

- В сотрудничестве с другими сторонами лаборатория отвечает за организацию испытания путем определения оцениваемых параметров качества, используемого материала для испытаний (в данный момент – критичного, но не чрезмерно), основы проведения испытания (например, методики, расстояний просмотра, организации сеансов, демонстрации последовательности испытательных элементов) и условий проведения испытания (например, условий просмотра, вступительного слова).
- Добровольно участвующим лабораториям предлагается предоставлять материал для испытаний в соответствии с соответствующими типовыми методами оценки параметра качества (на основе моделирования или аппаратного обеспечения).

- Другой партнер отвечает за монтаж ленты, используемой в испытании.
- Другие добровольно участвующие лаборатории проводят испытания с использованием предварительно смонтированной ленты. Испытание может быть испытанием вслепую. В этом случае лаборатория будет проводить испытание путем сбора экспертных оценок, необязательно зная, какие параметры качества оцениваются.
- Другому участнику, как правило, предлагается координировать сбор итоговых первоначальных данных с целью обработки и выпуска результатов, что может быть также осуществлено вслепую.
- Наконец, результаты интерпретируются на основе текстового/табличного или графического представления, и публикуется заключительный отчет.

Предлагаемый формат позволяет собирать результаты, получаемые в соответствии с процедурами испытаний, установленными на этапе определения испытания.

Формат соответствует методам оценки, описанным в части 1 и части 2 настоящей Рекомендации.

Он состоит из текстовых файлов, имеющих структуру, которая представлена в таблицах 1-4 и 1-5. Его синтаксис построен на текстовых данных и полях в дополнение к ограниченному набору зарезервированных знаков (например, "[", "]", " ", "┘" и "=").

Не существует существенного ограничения в плане объема (например, в отношении количества принимающих участие лабораторий, наблюдателей, испытательных последовательностей, границ шкалы голосования или типа периферийного устройства для выставления оценок).

ТАБЛИЦА 1-4

Формат текстового файла для определения результатов

Формат и синтаксис файла определения	Замечания
[рамки текста]┘ Тип = "DSCQS" или "DSIS II" и др. Количество сеансов = $1 \leq \text{целочисленное значение} \leq x$ ┘ Минимальное значение по шкале = <i>целочисленное значение</i> ┘ Максимальное значение по шкале = <i>целочисленное значение</i> ┘ Размер дисплея = <i>целочисленное значение</i> ┘ Марка и модель дисплея = <i>последовательность знаков</i> ┘ [РЕЗУЛЬТАТЫ]┘ Количество результатов = $1 \leq \text{целочисленное значение} \leq y$ ┘ Результат(j). Название(я) файла(ов) = <i>последовательность знаков.DAT</i> ┘ ... Результат(j). Название = <i>последовательность знаков</i> ┘ Результат(j). Лаборатория = <i>последовательность знаков</i> ┘ Результат(j). Число наблюдателей = $1 \leq \text{целочисленное значение} \leq N$ ┘ Результат(j). Обучение = "Да" или "Нет"┘	[Идентификатор раздела] Определение используемой методики по Рекомендации МСЭ-R ВТ.500 Количество сеансов ⁽¹⁾ , по которым было распределено испытание Установление шкалы (см. конкретные требования методики, если имеют место) Размер устройства отображения по диагонали (дюймы) [Идентификатор раздела] Количество рассматриваемых файлов с результатами ⁽¹⁾ Название файла Full.DAT (см. таблицу 7), включая путь доступа Название файла Custom Results (специализированные результаты) Идентификация лаборатории, проводящей испытание Общее число наблюдателей Указывает, включают ли собранные в ходе обучения оценки прикрепленный файл DAT

ТАБЛИЦА 1-4 (окончание)

Формат и синтаксис файла определения	Замечания
[Результат(<i>j</i>). Сеанс(<i>i</i>). Наблюдатели]↓ O(<i>k</i>). Имя = <i>последовательность знаков</i> .↓ O(<i>k</i>). Фамилия = <i>последовательность знаков</i> .↓ O(<i>k</i>). Пол = "Ж" или "М".↓ O(<i>k</i>). Возраст = <i>целочисленное значение</i> .↓ O(<i>k</i>). Профессия = <i>целочисленное значение</i> .↓ O(<i>k</i>). Расстояние = <i>целочисленное значение</i> .↓	[Идентификатор раздела] Идентификация наблюдателя Дополнительно Дополнительно Основные социально-экономические группы (например, рабочий, студент) Расстояние просмотра при высотах устройства отображения (например, 3 Н, 4 Н, 6 Н)

- (1) Сеанс: испытание может быть поделено на различные сеансы для удовлетворения требования максимальной продолжительности испытания. Те же или другие наблюдатели могут присутствовать на различных сеансах, в ходе которых им будет предложено оценить различные испытательные элементы. Объединение оценок, собранных на различных сеансах, дает полный набор результатов испытания (количество демонстраций × количество оценок на каждую демонстрацию). Результаты могут содержаться в различных файлах .DAT, которые будут предоставлены по каждой характеристике.

ТАБЛИЦА 1-5

Формат текстового файла первоначальных данных Результаты.DAT

Формат и синтаксис файла название файла.DAT	Замечания
целое целое целое.....↓	Файл DAT первоначальных данных состоит из значений оценок, разделенных пробелом. Для одного наблюдателя должна использоваться одна строка Первоначальные данные сохраняются в порядке ввода Данные могут быть распределены по различным файлам DAT, определенным в таблице 6, с помощью Результат(<i>j</i>). Название(я) файла(ов) ⁽¹⁾ .
целое целое целое.....↓	
целое целое целое.....↓	
...	

- (1) См. сноску ⁽¹⁾ к таблице 1-4.

Приложение 3 к части 1 (информационное)

Характеристики искажений содержания изображения

А3-1 Введение

После своего внедрения система будет использована для потенциально широкого диапазона программного материала, часть из которого она не сможет обработать без потери качества. Рассматривая пригодность системы, необходимо знать как долю программного материала, которая окажется критичной для данной системы, так и потери качества, которые можно в таких случаях ожидать. Фактически, для рассматриваемой системы требуется знать характеристики искажений содержания изображения.

Такая характеристика искажений особенно важна для систем, качество которых может уменьшаться неравномерно по мере того, как материал становится все более и более критичным. Например, некоторые цифровые и адаптивные системы могут поддерживать высокое качество в широком диапазоне программного материала, но ухудшать показатели за пределами этого диапазона.

А3-2 Получение характеристик искажения

В принципе, характеристика искажения содержания изображений устанавливает долю материала, вероятно, встречаемого при длительном воспроизведении, при котором система будет достигать некоторых уровней качества. Это проиллюстрировано на рисунке 1-5.

Характеристика искажения изображения может быть получена за четыре шага.

- *Шаг 1:* включает в себя определение алгоритмического измерения "критичности", с помощью которой можно было бы классифицировать ряд последовательностей изображений, которые были подвергнуты искажениям, присущим рассматриваемой системе или классу систем, так чтобы порядковый номер соответствовал бы оценке, которая была бы получена, если бы подобную задачу решали наблюдатели. Это измерение критичности может включать в себя аспекты моделирования зрительного восприятия.
- *Шаг 2:* включает в себя получение (при помощи измерения критичности большого количества отрывков типовых телевизионных программ) распределения оценок вероятности появления материала, имеющего различные уровни критичности для рассматриваемой системы или рассматриваемого класса систем. Пример такого распределения показан на рисунке 1-6.
- *Шаг 3:* включает в себя получение эмпирическим путем данных о способности системы поддерживать качество по мере увеличения критичности программного материала. Практически, это требует субъективной оценки качества, обеспечиваемого системой при работе с материалом, выбранного для дискретизации диапазона критичности, определенного на шаге 2. В результате этого получается функция, связывающая качество, обеспечиваемое системой с уровнем критичности программного материала. Пример такой функции показан на рисунке 1-7.
- *Шаг 4:* заключается в объединении информации, полученной на шагах 2 и 3 для того, чтобы получить характеристику искажений содержания изображения в форме, представленной на рисунке 1-5.

РИСУНОК 1-5

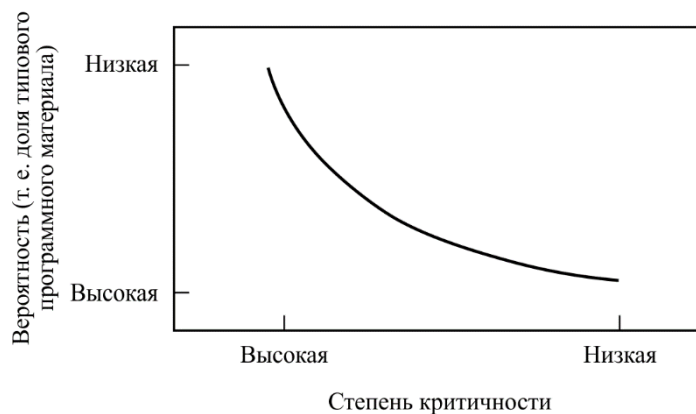
Графическое представление возможной характеристики искажения изображений



ВТ.0500-01-5

РИСУНОК 1-6

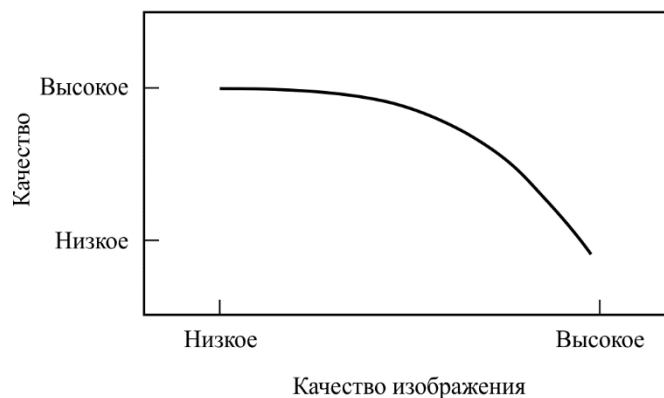
Вероятность появления материала с различными уровнями критичности



ВТ.0500-01-6

РИСУНОК 1-7

Возможная функция, связывающая качество, обеспечиваемое системой с критичностью программного материала



ВТ.0500-01-7

А3-3 Использование характеристики искажений

При получении общей картины качества, которое, вероятно, должно быть обеспечено в отношении всего возможного программного материала, характеристика искажений является важным средством рассмотрения пригодности систем. Характеристика искажений может быть использована тремя различными способами:

- для оптимизации параметров (исходного разрешения, скорости передачи битов, ширины полосы) системы на стадии разработки для того, чтобы она наиболее полно отвечала требованиям службы;
- для рассмотрения пригодности отдельной системы (то есть для того, чтобы предвидеть масштаб и серьезность искажений во время работы);
- для оценки относительной пригодности альтернативных систем (т. е. сравнения характеристик искажений и определения, какая из систем могла бы быть более пригодной для использования). Необходимо отметить, что, хотя различные системы подобного типа могут иметь одинаковый показатель критичности, возможно, что системы иного типа будут иметь другие показатели критичности. Однако так как характеристика искажений выражает только вероятность появления различных уровней качества, которые встречаются на практике, характеристики можно сравнивать непосредственно, даже если они получены на основе разных характерных для систем показателей критичности.

Хотя метод, описанный в данной Рекомендации, обеспечивает средства измерения характеристики искажения содержания изображения системы, он не может полностью предсказать приемлемость системы для зрителя телевизионной программы. Для получения такой информации может потребоваться, чтобы несколько зрителей просмотрели программы, закодированные с использованием системы, и затем изучить их замечания.

Пример характеристики искажения содержания изображения представлен в Приложении 1 к части 3.

Приложение 4 к части 1 (информационное)

Метод определения характеристики общих искажений в отношении содержания программы и условий передачи

A4-1 Введение

Характеристика общих искажений связывает воспринимаемое качество изображения с вероятностью появления на практике при непосредственном учете содержания программы и условий передачи.

В принципе, такая характеристика может быть получена исходя из результатов субъективного исследования, в котором задействованы ряд наблюдателей, продолжительности испытаний и точки приема, достаточные для получения выборки, которые представляли бы все возможные варианты содержания программ и условий приема. Однако на практике эксперимент такого вида невыполним.

В настоящем Дополнении описана альтернативная, более просто реализуемая процедура определения характеристики общих искажений. Этот метод содержит три этапа:

- анализ содержания программ;
- анализ каналов передачи; и
- получение характеристик общих искажений.

A4-2 Анализ содержания программ

Это этап состоит из двух операций. Во-первых, проводится соответствующая количественная оценка содержания программ и, во-вторых, оценивается вероятность появления этих измеренных значений на практике.

Количественная оценка содержания программ является статистической информацией, которая охватывает аспекты содержания программ, подчеркивающие способность рассматриваемой(ых) системы (систем) обеспечить ощущаемо верное воспроизведение программного материала. Ясно, что было бы очень полезно, чтобы это измерение было основано на приближенной модели восприятия. Однако при отсутствии такой модели может быть достаточной количественная оценка, касающаяся некоторых аспектов степени пространственного разнесения внутри кадров/полей и на протяжении их следования, при условии, что эта количественная оценка имеет примерно монотонную зависимость с воспринимаемым качеством изображения. Возможно, что для систем (или классов систем), основанных на фундаментально различных подходах к представлению изображения, необходимо будет использовать различные количественные оценки.

После того, как выбрана соответствующая количественная оценка, необходимо оценить вероятности, с которыми появляются возможные статистические значения. Это может быть сделано двумя способами:

- при помощи эмпирической процедуры анализируется случайная выборка, возможно, двухсот 10-секундных сегментов программ студийного формата с подходящим разрешением, скоростью следования полей и форматом изображения для рассматриваемой(ых) системы (систем). В результате анализа этой выборки будут получены относительные частоты появления статистических значений, которые принимаются в качестве значений оценки вероятности появления на практике; или

- при помощи теоретического метода: для оценки вероятностей используется теоретическая модель. Необходимо отметить, что, хотя эмпирический метод является предпочтительным, в конкретных случаях может возникнуть необходимость использования теоретического метода (например, когда нет достаточной информации о содержании программ, в частности, при появлении новых технологий производства).

В результате упомянутого выше анализа будет получено распределение вероятности для значений статистики содержания программ (см. также Приложение 3). Этот результат будет объединен с результатами анализа условий передачи для подготовки к завершающему этапу процесса.

A4-3 Анализ канала передачи

Этот этап также включает в себя две операции. В первой операции проводится измерение качества канала передачи. И во второй операции оцениваются вероятности, с которыми значения, полученные в результате измерения, появляются на практике.

Измерение канала передачи – это статистика, которая охватывает аспекты качества канала, влияющие на способность рассматриваемой(ых) системы (систем) обеспечивать ощутимо верное воспроизведение исходного материала. Ясно, что было бы очень полезно основывать это измерение на приближенной модели восприятия. Однако, при отсутствии такой модели, могут быть достаточными измерения, касающиеся некоторого аспекта ограничения, налагаемого каналом, при условии, что это измерение имеет закономерную связь с воспринимаемым качеством изображения. Возможно, что для систем (или классов систем), основанных на фундаментально различных подходах к каналному кодированию, необходимо будет использовать различные измерения.

После того, как выбрано соответствующее измерение, необходимо оценить вероятности, с которыми появляются возможные статистические значения. Это может быть сделано двумя способами:

- при помощи эмпирической процедуры: качество канала измеряется для, возможно, 200 случайно выбранных моментов времени и точек приема. В результате анализа этой выборки будут получены относительные частоты появления статистических значений, которые принимаются в качестве значения оценки вероятности появления на практике; или
- при помощи теоретического метода: для оценки вероятностей используется теоретическая модель. Необходимо отметить, что хотя эмпирический метод является предпочтительным, в конкретных случаях может возникнуть необходимость использования теоретического метода (например, когда нет достаточной информации о качестве канала, в частности, при появлении новых технологий производства).

В результате упомянутого выше анализа будет получено распределение вероятности для значений статистики канала. Этот результат будет объединен с результатами анализа содержания программ для подготовки к завершающему этапу процесса.

A4-4 Получение характеристики общих искажений

Этот этап включает в себя субъективный эксперимент, в котором содержание программ и условия передачи изменяются совместно, в соответствии с вероятностями, полученными на первых двух этапах.

Основной используемый метод – процедура непрерывной оценки качества с двумя источниками воздействия, и, в частности, для движущихся последовательностей рекомендуется 10-секундный вариант (см. Приложение 2 к части 2). В этом случае эталоном является изображение студийного качества соответствующего формата (например, с разрешением, скоростью кадров и форматом изображения, соответствующими рассматриваемой системе (системам)). В отличие от этого испытательный сигнал представляет собой то же самое изображение, которое было принято в рассматриваемой системе (системах) при выбранных условиях работы канала.

Материал для испытаний и условия работы канала выбираются в соответствии с вероятностями, полученными на первых двух этапах метода. Фрагменты материала для испытания, каждый из которых был предварительно проанализирован с целью определения преобладающего значения, соответствующего статистике содержания программ, включаются в отобранный набор. Затем материал

выбирается из этого набора таким образом, чтобы он охватывал весь диапазон возможных статистических значений, реже – с менее критическими уровнями и более часто – с более критическими уровнями. Возможные значения статистики канала выбираются подобным путем. Затем эти два независимых источника влияния объединяются случайным образом для того, чтобы получить в результате сочетание содержания и условий работы канала с известной вероятностью.

Результаты таких исследований, которые связывают воспринимаемое качество изображения с вероятностью появления на практике, используются затем для рассмотрения пригодности системы или для сравнения систем на предмет пригодности.

Приложение 5 к части 1 (информационное)

Влияние контекста

Влияния контекста возникают, когда на субъективную классификацию изображения воздействуют порядок и серьезность демонстрируемых искажений. Например, если сильно искаженное изображение демонстрируется после последовательности слабо искаженных изображений, то зрители невольно могут дать этому изображению более низкую оценку, по сравнению с той, которую они могли бы дать в обычной ситуации.

Группа из четырех лабораторий, расположенных в различных странах, изучала возможные влияния контекста, связанные с результатами, полученными при использовании трех методов (метода DSCQS, варианта II метода DSIS и метода сравнения) оценки качества изображения. Материал для испытаний был создан с применением кодирования MPEG (ML и MP) и снижения разрешения по горизонтали. К каждой серии испытаний (одна серия, описывающая слабые контекстные искажения, и другая серия, описывающая сильные искажения) применялись четыре основных условия испытаний (B1, B2, B3, B4) и шесть контекстных условий испытаний. К обеим сериям применялись три метода испытаний. Влияния контекста – это различия между результатами в отношении испытания, содержащего, главным образом, слабые искажения, и испытанием, содержащим, главным образом, сильные искажения. Базовые условия испытаний B2 и B3 использовались для определения влияний контекста.

Результаты, полученные всеми лабораториями, указывают на отсутствие влияний контекста в случае метода DSCQS. В случаях метода DSIS и метода сравнения влияния контекста были очевидными, а наиболее сильное влияние было обнаружено в случае варианта II метода DSIS. Результаты указывают, что преимущественно слабые искажения могут обуславливать недооценку изображения, а преимущественно сильные искажения – его переоценку.

Результаты исследования показывают, что метод DSCQS – это лучший метод для сведения к минимуму влияний контекста при субъективной оценке качества изображения, рекомендованной МСЭ-R.

Более подробная информация об упомянутом выше исследовании содержится в Отчете МСЭ-R ВТ.1082.

Приложение 6 к части 1 (информационное)

Измерения пространственной и временной информации

Представленные ниже измерения пространственной и временной информации имеют единственное значение для каждого кадра полной испытательной последовательности. В результате получается временной ряд значений, обычно различающихся в некоторой степени. Представленные ниже измерения информации о восприятии устраняют эту изменчивость с помощью максимальной функции

(максимальное значение для последовательности). Сама изменчивость может быть с успехом изучена, например с помощью графиков пространственно-временной информации, построенных на кадровой основе. Использование распределений информации по испытательной последовательности позволяет также лучше оценить сцены, содержащие смонтированные сцены.

Информация о пространственном восприятии (SI): Измерение, которое обычно указывает объем пространственных деталей в изображении. Как правило, он больше для более сложных в пространственном отношении сцен. Эта величина не является измерением энтропии и не связана с информацией, определенной в теории связи. Информация о пространственном восприятии SI основана на фильтре Собеля. Каждый кадр видеоизображения (плоскость яркости) в момент n (F_n) прежде всего фильтруется фильтром Собеля [Sobel (F_n)]. Затем рассчитывается стандартное отклонение по пикселям (std_{space}) в каждом кадре, прошедшем через фильтр Собеля. Данная операция повторяется для каждого кадра последовательности видеоизображения, и в результате получается временной ряд пространственной информации сцены. Для представления содержания пространственной информации сцены из временного ряда выбирается максимальное значение (max_{time}). Этот процесс может быть представлен в форме уравнения:

$$SI = \max_{time} \{std_{space} [Sobel(F_n)]\}.$$

Информация о временном восприятии (TI): Измерение, которое обычно указывает объем временных изменений в последовательности видеоизображения. Как правило, он выше для последовательностей с быстро движущимися изображениями. Эта величина не является измерением энтропии и не связана с информацией, определенной в теории связи.

Измерение временной информации TI рассчитывается как максимальное значение по времени (max_{time}) стандартного отклонения в пространстве (std_{space}) величины $M_n(i, j)$ по всем i и j .

$$TI = \max_{time} \{std_{space} [M_n(i, j)]\},$$

где $M_n(i, j)$ – разница между пикселями в одной и той же точке кадра, принадлежащими двум последовательным кадрам:

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j),$$

где $F_n(i, j)$ – пиксел в i -й строке и j -й колонке n -го кадра во времени.

ПРИМЕЧАНИЕ. – Для сцен, содержащих смонтированные сцены, могут предоставляться два значения: одно – в случае включения смонтированной сцены в измерение временной информации, другое – в случае ее исключения из измерения.

Приложение 7 к части 1 (информационное)

Термины и определения

Algorithm	One or several image processing operations	Алгоритм	Одна или несколько операций по обработке изображения
AVI	Audio video interleaved		Перемежение аудио- и видеосигналов
CCD	Charge coupled device		Прибор с зарядовой связью
CI	Confidence interval		Доверительный интервал
CIF	Common intermediate format (defined in Recommendation H.261 for video phone: 352 lines \times 288 pixels)		Единый промежуточный формат (определен для видеофонов в Рекомендации H.261: 352 строки \times 288 пикселей)
CRT	Cathode ray tube		Телевизионная трубка

DSCQS	Double stimulus using a continuous quality scale method		Метод двух источников воздействия с непрерывной шкалой качества
DSIS	Double stimulus using an impairment scale method		Метод двойного входного сигнала с использованием шкалы ухудшения
LCD	Liquid crystal display	УОЖК	Устройство отображения на жидких кристаллах
MOS	Mean opinion score		Средняя экспертная оценка
SC	Stimulus comparison method		Методы сравнения входных сигналов
PDP	Plasma display panel	ПУО	Плазменное устройство отображения
PS	Programme segment	СП	Сегмент программы
QCIF	Quarter CIF (defined in Recommendation H.261 for video phone: 176 lines × 144 pixels)		Четверть единого промежуточного формата (определен для видеофонов в Рекомендации H.261: 176 строк × 144 пикселей)
SAMVIQ	Subjective assessment of multimedia video quality		Субъективная оценка качества мультимедийного видеоизображения
Sequence	Scene with combined processing or without processing	Последовательность	Сцена, смешанная с обработанными частями или без обработки
Scene	Audio-visual content	Сцена	Аудиовизуальный контент
S/N	Signal-to-noise ratio		Отношение сигнал/шум
SI	Spatial information		Пространственная информация
SIF	Standard intermediate format [picture formats defined in ISO 11172 (MPEG-1): 352 lines × 288 pixels × 25 frames/s and 352 lines × 240 pixels × 30 frames/s]		Стандартный промежуточный формат [определен в стандарте ИСО 11172 (MPEG-1): 352 строки × 288 пикселей × 25 кадров/с и 352 строки × 240 пикселей × 30 кадров/с]
SP	Simultaneous presentation	ОД	Одновременная демонстрация
SQCIF	Sub-QCIF		суб-QCIF
SS	Single stimulus method		Метод с использованием одного входного сигнала
SSCQE	Single stimulus using a continuous quality evaluation method		Метод непрерывной оценки качества при одном источнике воздействия
std	Standard deviation		Стандартное отклонение
TI	Temporal information	ВИ	Временная информация
TP	Test presentation	ТД	Испытательная демонстрация
TS	Test session	СИ	Сеанс испытания
VTR	Video tape recorder	ВМ	Видеомагнитофон

ЧАСТЬ 2

Описание методик субъективной оценки изображения

1 Введение

В этой части подробно описана каждая из методик оценки изображения, необходимых для субъективных оценок качества изображения. В некоторых случаях имеются отличия от общих особенностей оценки, описанных в разделе 2 части 1.

Для корректной интерпретации результатов субъективных оценок качества изображения в других лабораториях важно дать подробные пояснения к процедурам и зафиксировать все изменения в используемой методике, представив всю дополнительную информацию, которая может потребоваться другой лаборатории, желающей повторить процедуру оценки.

2 Рекомендуемые методики оценки изображения

Приложение 1	Метод с двумя источниками воздействия и с использованием шкалы искажений (DSIS)
Приложение 2	Метод двух источников воздействия с непрерывной шкалой качества (DSCQS)
Приложение 3	Методы с одним источником воздействия (SS)
Приложение 4	Методы сравнения воздействий
Приложение 5	Непрерывная оценка качества при одном источнике воздействия (SSCQE)
Приложение 6	Метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE)
Приложение 7	Субъективная оценка качества мультимедийного видеоизображения (SAMVIQ)
Приложение 8	Протокол просмотра экспертами (EVP) для оценки качества видеоматериала

3 Замечания

Другие методы, такие как методы многомерного масштабирования и многовариантные методы, описаны в Отчете МСЭ-R ВТ.1082 и все еще изучаются.

Всем описанным ранее методам присущи и сильные стороны, и ограничения, и невозможно определенно рекомендовать один из них. Следовательно, выбор метода, наиболее подходящего к конкретным обстоятельствам, остается на усмотрение исследователя.

Ограничения, присущие различным методам, наводят на мысль, что может быть неразумным слишком полагаться на один метод.

Следовательно, может быть целесообразным рассмотреть более "полные" подходы, такие как использование нескольких методов или использование многомерного подхода.

Приложение 1 к части 2

Метод с двумя источниками воздействия и с использованием шкалы искажений (DSIS) ("метод ЕСР")

A1-1 Общее описание

Типичный метод оценки может быть применен для оценки либо новой системы, либо искажений, вносимых на пути передачи. Первые шаги для организации испытаний должны включать подбор испытательного материала, достаточного для получения представительной оценки, и установлении того, в каких условиях следует проводить испытания. Если представляет интерес изменение влияния параметра, то необходимо выбрать такой набор его значений, который перекроет весь диапазон изменения искажений с использованием небольшого числа больших шагов равного размера. Если оценивается новая система, для которой значения параметров не могут быть изменены таким способом, тогда либо необходимо добавить субъективные искажения, либо должен быть использован иной метод, например описанный в Приложении 2 к настоящей части 2.

Метод с двумя источниками воздействия и с использованием шкалы искажений (DSIS) (метод ЕСР) – циклический, в нем оценщику сначала демонстрируется неискаженный эталон, а затем – то же самое изображение с искажениями. После этого его просят оценить второй образец относительно первого. Во время сеанса, который длится до получаса, оценщику демонстрируют наборы изображений или их последовательности в случайном порядке и со случайными искажениями, охватывая все возможные комбинации. Неискаженное изображение также включается в последовательность, которая должна быть оценена. В конце серии испытаний вычисляют среднюю оценку для каждого условия испытаний и испытательного изображения.

Метод использует шкалу искажений, для которой обычно стабильность результатов выше при небольших искажениях, чем при значительных. Хотя этот метод иногда используется с искажениями, изменяющимися в ограниченном диапазоне, правильнее использовать его с полным диапазоном искажений.

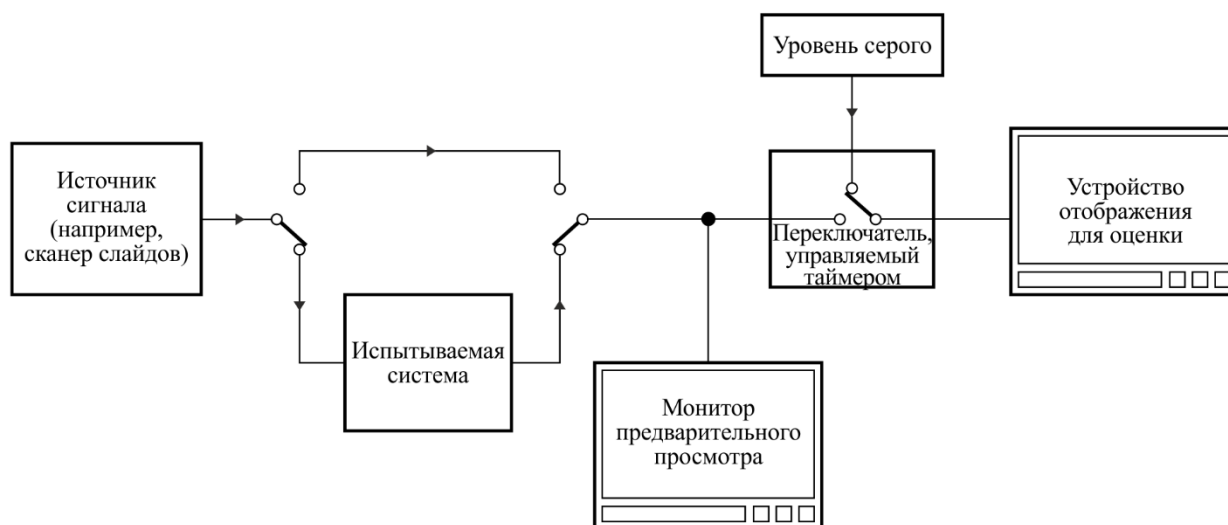
A1-2 Общий порядок

Условия просмотра, исходные сигналы, материал для испытаний, наблюдатели и представление результатов определяются или выбираются в соответствии с разделом 2 части 1.

Общий набор оборудования для испытательной установки должен быть таким, как показано на рисунке 2-1.

РИСУНОК 2-1

Общая компоновка испытательной системы для метода DSIS



ВТ.0500-02-1

Оценщики смотрят на устройство отображения для оценки, сигнал на которое подается через переключатель, управляемый таймером. Сигнал на этот переключатель может подаваться непосредственно от источника сигнала или косвенно через испытываемую систему. Оценщикам представляют серию испытательных изображений или последовательностей. Они соединены попарно, причем первое изображение в паре проходит по прямому пути от источника, а второе представляет собой то же самое изображение, прошедшее через испытываемую систему.

A1-3 Представление материала для испытаний

Сеанс испытаний включает в себя ряд представлений. Существует два варианта структуры представления – I и II, описанные ниже.

Вариант I: Эталонное изображение или последовательность и оцениваемое изображение или последовательность представляются только один раз, как это показано на рисунке 2-2 (а).

Вариант II: Эталонное изображение или последовательность и оцениваемое изображение или последовательность представляются дважды, как это показано на рисунке 2-2 (б).

Вариант II, который занимает больше времени, чем вариант I, может быть использован, если требуется различить очень небольшие искажения или испытывается последовательность движущихся изображений.

A1-4 Шкалы оценок

Следует использовать пятибалльную шкалу оценки искажений:

- 5 незаметно;
- 4 заметно, но не раздражает;
- 3 слегка раздражает;
- 2 раздражает;
- 1 очень раздражает.

Оценщики должны использовать бланк, содержащий подробную шкалу и пронумерованные графы или иные средства подобного рода для записи своих оценок.

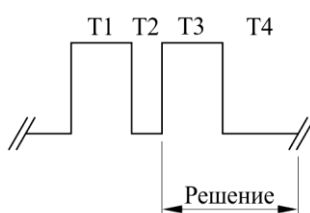
A1-5 Представление оценок

В начале каждого сеанса наблюдателям дается объяснение о типе испытаний, шкале оценок, последовательности и порядке следования (эталонное изображение, серый, оцениваемое изображение, время для принятия решения). Диапазон и тип оцениваемых искажений следует проиллюстрировать на примерах изображений, отличных от тех, которые будут использованы в испытаниях, но сравнимых по восприятию. Необязательно наихудшему качеству будет соответствовать самая маленькая субъективная оценка. Необходимо попросить наблюдателей опираться при оценке на общее впечатление, производимое изображением, и выражать свое решение терминами формулировок, используемых для установления субъективной шкалы.

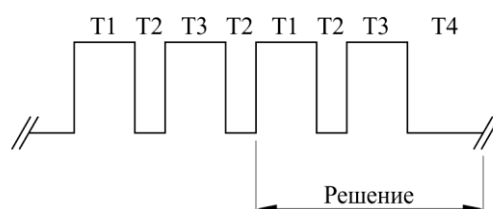
Наблюдателей следует просить рассматривать изображение в течение полных периодов T1 и T3. Принятие решения должно быть разрешено только в течение T4.

РИСУНОК 2-2

Структура представления материала для испытаний



а) Вариант I



б) Вариант II

ВТ.0500-02-2

Этапы представления

T1 = 10 с	Эталонное изображение
T2 = 3 с	Средний серый, создаваемый видеосигналом с уровнем примерно 200 мВ
T3 = 10 с	Испытательный сигнал
T4 = 5–11 с	Средний серый

Опыт показывает, что увеличение периодов T1 и T3 свыше 10 с не повышает способности оценщиков давать правильную оценку последовательностям изображений.

A1-6 Сеанс испытаний

Изображения и искажения должны быть представлены в псевдослучайном порядке и, предпочтительно, в различной последовательности для каждого сеанса. Во всяком случае, одни и те же испытательные изображения или последовательности никогда не должны быть представлены в одном и том же порядке с теми же или различными уровнями искажений.

Диапазон искажений должен быть выбран таким образом, чтобы большинством наблюдателей использовались все оценки; средняя оценка (усредненная по всем оценкам, принятым в ходе эксперимента) должна быть около трех.

Сеанс не должен длиться более примерно получаса, включая объяснения и предварительные этапы; испытательные последовательности могут начинаться с нескольких изображений, указывающих диапазон искажений; оценки, поставленные этим изображениям, могут не приниматься в расчет при подведении окончательных итогов.

Дополнительные соображения по выбору уровней искажений приведены в Приложении 2 к части 1.

Приложение 2 к части 2

Метод двух источников воздействия с непрерывной шкалой качества (DSCQS)

A2-1 Общее описание

Типичный метод может быть применен для оценки либо новой системы, либо влияний трактов передачи на качество. Идея метода с двумя источниками воздействия особенно полезна в тех случаях, когда в отношении источников воздействия невозможно обеспечить такие условия испытания, которые представляют полный диапазон качества.

Данный метод является циклическим, при котором оценщика просят рассмотреть пару изображений, полученных от одного и того же источника, но одно из них при этом подверглось рассматриваемому процессу обработки, а второе получено непосредственно от источника. Его просят оценить качество обоих изображений.

Во время сеансов, которые длятся до получаса, оценщику представляют пары изображений (случайно подобранные) в случайном порядке, со случайными искажениями, но при этом должны быть представлены все требуемые сочетания. В конце сеансов вычисляются средняя оценка для каждого условия испытаний и испытательного изображения.

A2-2 Общий порядок

Условия просмотра, исходные сигналы, материал для испытаний, наблюдатели и ознакомление с оценкой определяются или выбираются в соответствии с разделом 2 части 1. Сеанс испытаний описан в разделе A1-6 Приложения 1 к части 2.

Общий набор оборудования для испытательной системы должен быть таким, как показано на рисунке 2-3.

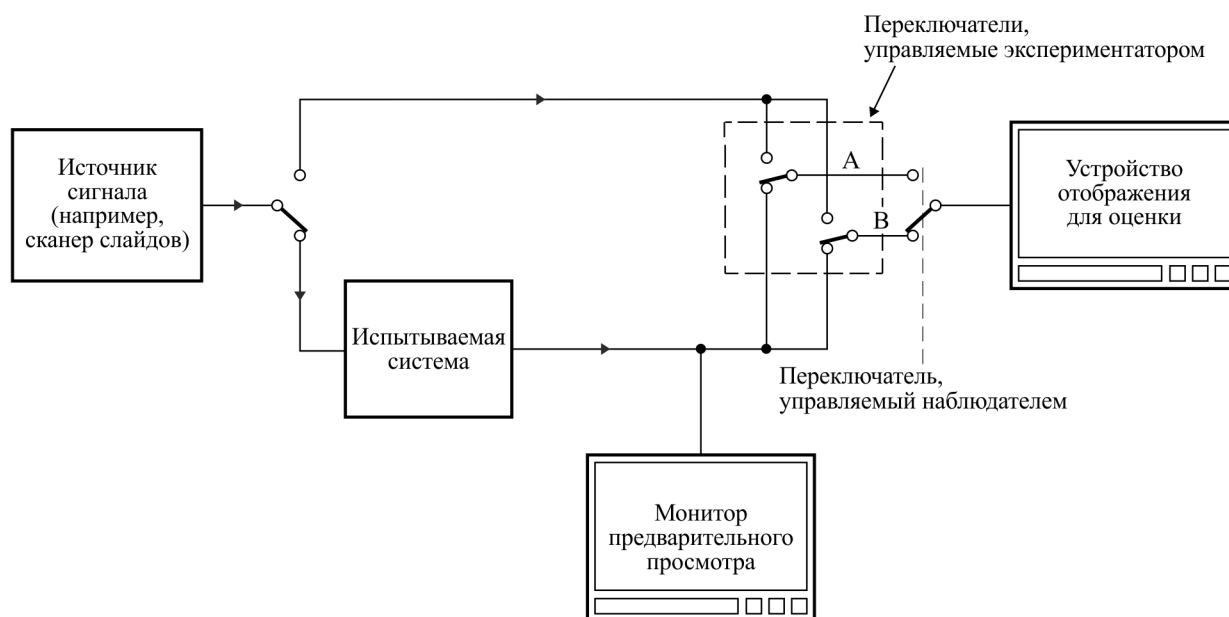
A2-3 Представление материала для испытаний

Сеанс испытаний состоит из ряда представлений. В случае варианта I, при котором участвует один наблюдатель, во время каждого представления оценщик может просматривать сигналы А и В до тех пор, пока он не оценит качество каждого сигнала. Оценщик обычно может сделать это два или три раза за период 10 секунд. В случае варианта II, в котором одновременно участвуют несколько наблюдателей, перед регистрацией результатов пара изображений демонстрируется один или несколько раз в течение равных промежутков времени для того, чтобы оценщик мог мысленно оценить их качества; затем эта пара демонстрируется еще один или несколько раз, при этом результаты регистрируются. Число повторений зависит от длительности испытательных последовательностей. Для неподвижных изображений могут быть целесообразными 3–4-секундные последовательности и пять повторений (принятие решения в течение последних двух). Для подвижных изображений с изменяемыми во времени артефактами может быть целесообразной последовательность длительностью 10 секунд с двумя повторениями (принятие решения во время второго). Структура представлений показана на рисунке 2-4.

В случае, когда по практическим соображениям длительность имеющихся последовательностей ограничена менее чем 10 секундами, эти короткие последовательности могут быть использованы в композициях как сегменты до достижения времени показа, равного 10 секундам. Для сведения к минимуму разрыва в местах соединения следующие друг за другом сегменты могут быть инвертированы во времени (такой способ иногда называют "палиндромным" показом). Необходимо принять меры для того, чтобы условия испытаний при демонстрации инвертированных во времени отрезков сигнала представляли собой случайные процессы, это может быть достигнуто путем передачи инвертированного во времени сигнала источника через испытываемую систему.

РИСУНОК 2-3

Общая компоновка испытательной системы для метода DSCOS



ВТ.0500-02-3

Существует два варианта данного метода, в общих чертах описанных ниже.

- Вариант I** Наблюдатель, который обычно находится в одиночестве, имеет возможность переключать два изображения A и B до тех пор, пока не убедится в том, что его мнение о каждом изображении устойчиво. По линиям A и B подают прямое эталонное изображение, либо изображение, прошедшее через испытываемую систему, но то, какое из них по какой линии подается, изменяется случайным образом в зависимости от условий испытания. Эти условия отмечают экспериментатором, но не известны наблюдателю.
- Вариант II** Наблюдателям показывают последовательно изображения, полученные с линий A и B для того, чтобы у них сложилось мнение относительно каждого изображения. Изображения на линии A и B для каждого представления подают так же, как было описано выше для варианта I.

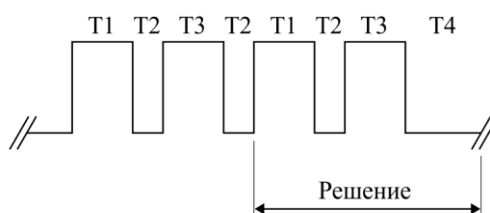
A2-4 Шкала оценок

Данный метод требует оценки обоих вариантов каждого испытательного изображения. В каждой паре одно испытательное изображение является неискаженным, тогда как другое представление может либо содержать, либо не содержать искажений. Неискаженное изображение включается для того, чтобы служить эталоном, но наблюдателям не говорят о том, какое из изображений эталонное. В серии испытаний позиция эталонного изображения выбирается псевдослучайным образом.

Наблюдателям предлагают просто оценить общее качество каждого представленного изображения, проставив отметку на вертикальной шкале. Такие вертикальные шкалы печатаются попарно для сдвоенного представления каждого испытательного изображения. Эти шкалы представляют собой систему непрерывной оценки во избежание ошибок квантования, но они разделены на пять отрезков равной длины, соответствующих обычной пятибалльной шкале качества МСЭ-R. Термины, определяющие различные уровни, являются теми же, что используются обычно; но они включены для общего руководства и напечатаны только слева от первой шкалы в каждом ряду из десяти сдвоенных столбцов на бланке оценок. На рисунке А2-3 показана часть типичного бланка оценок. Для того чтобы исключить любую возможность путаницы между делениями шкалы и отметками результатов испытаний, шкалы печатаются синим цветом, а запись результатов ведется черным цветом.

РИСУНОК 2-4

Структура представления материала для испытания



ВТ.0500-02-4

Этапы представления

T1 = 10 с	Испытательная последовательность А
T2 = 3 с	Средний серый, уровень видеосигнала примерно 200 мВ
T3 = 10 с	Испытательная последовательность В
T4 = 5–11 с	Средний серый

РИСУНОК 2-5

Часть формы для оценки качества с использованием непрерывных шкал*

	27		28		29		30		31	
	А	В	А	В	А	В	А	В	А	В
Отлично										
Хорошо										
Удовлетворительно										
Плохо										
Неприемлемо										

ВТ.0500-02-5

* При планировании порядка представления материала для испытаний в случае метода DSCQS желательно, чтобы экспериментатор включал проверки для обеспечения уверенности в том, что эксперимент свободен от систематических ошибок. Однако метод выполнения таких проверок, дающих уверенность, еще изучается.

A2-5 Анализ результатов

Пары оценок (эталонного и испытываемого изображений) для каждого условия испытаний преобразуют из измерений длины на бланке оценок в нормированные оценки, расположенные в диапазоне от 0 до 100. Затем рассчитываются разности между оценкой в условиях эталонного и испытываемого изображения. Более подробно процедура описана в Приложении 2 к части 1.

Опыт показал, что оценки, полученные для различных испытательных последовательностей, зависят от критичности используемого для испытаний материала. Более полное понимание качества кодека может быть достигнуто путем отдельного представления результатов для различных испытательных последовательностей, а не только в форме суммарных средних значений по всем испытательным последовательностям, используемым в оценке.

Если результаты для отдельных испытательных последовательностей выстроены по абсциссе в порядке ранга критичности испытательной последовательности, то можно представить приближенное графическое описание характеристики нарушения содержания изображения в испытываемой системе. Однако эта форма представления описывает только качество кодека и не предоставляет указания вероятности появления последовательностей с заданной степенью критичности (см. Приложение 2 к части 1). Необходимо провести дополнительные исследования, касающиеся критичности испытательных последовательностей и вероятности появления последовательностей с заданным уровнем критичности, до возможного получения более полного представления о качестве системы.

A2-6 Интерпретация результатов

При использовании данного метода DSCQS может быть рискованно и даже ошибочно делать выводы о качестве исследуемых изображений путем увязки цифровых значений DSCQS с характеристиками, получаемыми на основе других протоколов испытаний (например, "незаметно"; "заметно, но не раздражает"; ... полученными на основе метода DSIS).

Отмечается, что результаты, полученные методом DSCQS, должны рассматриваться не как абсолютные оценки, а как разницы в оценках в условиях эталонного и испытательного изображений. Таким образом, ошибочно увязывать оценки с единственным описанием качества, даже если они получены с помощью самого протокола DSCQS (например, "отлично", "хорошо", "удовлетворительно"...).

При любой процедуре испытаний важно принять решение в отношении критериев приемлемости до начала проведения оценки. Это особенно важно при применении метода DSCQS, поскольку у неопытных пользователей существует тенденция к неправильному пониманию значения величин на шкале качества, получаемых с помощью этого метода.

Приложение 3 к части 2

Методы с одним источником воздействия (SS)

В методах с одним источником воздействия демонстрируется одно изображение или последовательность изображений, и оценщик ставит балл в отношении всей демонстрации. Материал для испытания может включать только испытательные последовательности или испытательные последовательности и соответствующую им эталонную последовательность. В последнем случае эталонная последовательность представляется как отдельное воздействие, которое оценивается как любое другое испытательное воздействие.

A3-1 Общий порядок

Условия просмотра, сигналы источника, диапазон условий и привязки, наблюдатели, ознакомление с оценкой и представление результатов определены или выбраны в соответствии с разделом 2 части 1.

А3-2 Выбор материала для испытаний

Содержание испытательных изображений для лабораторных испытаний должно выбираться, как описано в пункте 2.3 части 1.

После выбора содержания испытательные изображения подготавливают таким образом, чтобы они отражали рассматриваемые варианты исследования или диапазон(ы) воздействия одного (или нескольких) факторов. Если изучается воздействие двух или более факторов, то изображения могут быть подготовлены двумя способами. В первом случае каждое изображение представляет один уровень воздействия только одного фактора. В другом случае каждое изображение представляет один уровень воздействия всех рассматриваемых факторов, но в ходе показа каждый уровень воздействия каждого фактора появляется при каждом уровне всех других факторов. Оба метода позволяют получить четкие результаты о влиянии отдельных факторов. Последний метод позволяет также обнаружить взаимодействие между факторами (то есть неаддитивные влияния).

А3-3 Сеанс испытаний

Сеанс испытаний состоит из серии экспериментов по оценке. Они должны быть представлены в случайном порядке и, предпочтительно, в различной для каждого наблюдателя случайной последовательности. При использовании одного случайного порядка последовательностей существует два варианта структуры представлений I (SS) и II (одно многократно повторяемое воздействие (SSMR)), приведенные ниже:

- a) Испытательные изображения или последовательности представляются только один раз в течение сеанса испытаний; в начале первых сеансов должны быть представлены некоторые тренировочные последовательности (как описано в пункте 2.7 части 1); обычно эксперимент обеспечивает, чтобы то же изображение не демонстрировалось дважды подряд при том же уровне искажений.

Типичный эксперимент по оценке включает три показа: средне серое поле адаптации, воздействие и средне серое постэкспозиционное поле. Длительность показа варьируется в зависимости от задачи просмотра, материалов и рассматриваемых мнений или факторов, но нередко составляет 3, 10 и 10 секунд соответственно. Балл или баллы наблюдателя могут быть выставлены в ходе показа воздействия или постэкспозиционного поля.

- b) Испытательные изображения или последовательности демонстрируются три раза, в результате чего организуется сеанс испытаний из трех демонстраций, каждая из которых включает все изображения или последовательности, испытываемые только один раз; сообщение на дисплее объявляет о начале каждой демонстрации (например, демонстрация 1); первая демонстрация используется для укрепления мнения наблюдателя; данные, полученные на основе этой демонстрации не должны учитываться в результатах испытания; оценки, присвоенные изображениям или последовательностям получают путем усреднения данных второй и третьей демонстраций; обычно в эксперименте обеспечивается применение следующих ограничений к случайному порядку изображений или последовательностей в рамках каждой демонстрации:

- расположение данного изображения или данной последовательности не является тем же в других демонстрациях;
- данное изображение или данная последовательность не располагаются непосредственно перед тем же изображением или той же последовательностью в других демонстрациях.

Типичный эксперимент по оценке включает два показа: воздействие и средне серое постэкспозиционное поле. Длительность показов варьируется в зависимости от задачи просмотра, материалов и рассматриваемых мнений или факторов, но нередко составляет 10 и 5 секунд соответственно. Балл или баллы зрителя должны быть выставлены в ходе показа только постэкспозиционного поля.

В случае варианта II (SSMR) время, требуемое для осуществления сеанса испытаний, явно увеличивается (45 секунд вместо 23 секунд для каждого испытываемого изображения или испытываемой последовательности); однако в случае данного варианта существенно уменьшается

зависимость результатов, получаемых при варианте I, от порядка следования изображений или последовательностей в рамках сеанса.

Кроме того, экспериментальные результаты показывают, что вариант II позволяет охватить около 20% в рамках диапазона оценок.

A3-4 Типы методов с одним источником воздействия (SS)

Обычно в телевизионных оценках используется три типа методов с одним источником воздействия.

A3-4.1 Методы оценок с использованием определенных категорий

В случае оценок с использованием определенных категорий наблюдатели присваивают изображению или последовательности изображений одну категорию из предлагаемого набора, которые обычно определены в семантических терминах. Категории могут отражать решения о том, обнаружен или нет искомый атрибут (например, для установления порога для искажений). Чаще используются шкалы категорий, которые определяют качество изображения и искажения изображения, и шкалы МСЭ-R, приведенные в таблице 2-1. При эксплуатационном контроле иногда используются промежуточные оценки. В особых случаях применяются шкалы для определения разборчивости текста, усилий, которые необходимо приложить для того, чтобы его прочесть, и полезности изображений.

ТАБЛИЦА 2-1

Шкалы МСЭ-R для оценки качества и искажений

Пятибалльная шкала			
Качество		Искажения	
5	Отлично	5	Незаметно
4	Хорошо	4	Заметно, но не раздражает
3	Удовлетворительно	3	Слегка раздражает
2	Плохо	2	Раздражает
1	Неприемлемо	1	Очень раздражает

Этот метод дает распределение мнений по шкале категорий для каждого из условий. Способ анализа ответов зависит от оценки (обнаружение и т. д.) и от искомой информации (определение порога, баллов или основной тенденции условий, психологических "дистанций" между различными условиями). Существует много методов анализа.

A3-4.2 Численные методы оценок с использованием категорий

Процедура с одним источником воздействия, в которой используется 11-ступенчатая цифровая шкала категорий (SSNSC), была исследована и сравнена с графической и относительной шкалами. Это исследование, описанное в Отчете МСЭ-R ВТ.1082, показывает явное преимущество метода SSNSC по чувствительности и стабильности в тех случаях, когда нет эталонных сигналов.

A3-4.3 Методы оценок без использования категорий

При оценивании без использования категорий наблюдатели присваивают значение каждому показанному изображению или последовательности изображений. Существует два варианта такого метода.

При использовании непрерывной шкалы – варианта метода с использованием категорий – оценщик присваивает каждому изображению или последовательности изображений точку на линии, проведенной между двумя семантическими метками (то есть концами шкалы категорий, приведенной в таблице 3). Для справки, шкала может включать дополнительные метки в промежуточных точках. Расстояние от конца шкалы принимается в качестве оценки каждого условия.

При использовании цифровой шкалы оценщик присваивает каждому изображению или последовательности изображений цифру, которая отражает уровень его оценки в некоторой области (например, четкость изображения). Диапазон используемых цифр может быть как ограниченным (например, от 0 до 100), так и неограниченным. Иногда присвоенная цифра описывает оцениваемый уровень в абсолютном выражении (без прямой ссылки на уровень какого-либо иного изображения или последовательности изображений), как при некоторых формах оценки амплитуды. В иных случаях цифра описывает оцениваемый уровень по отношению к предварительно просмотренному "стандарту" (например, оценка амплитуды, подробности изображения и относительная оценка).

Обе формы дают в результате распределение цифровых оценок для каждого из условий. Используемый метод анализа зависит от типа суждения и требуемой информации (например, баллов, основной тенденции, психологических "дистанций").

A3-4.4 Методы оценки качества

Некоторые аспекты обычного просмотра могут быть выражены в отношении качества выполнения поручаемых извне задач (поиск целевой информации, чтение текста, идентификация объектов и т. д.). Затем измеренный показатель работы, например точность или скорость, с которой такие задачи выполняются, может быть использован в качестве оценки изображения или последовательности изображений.

Методы оценки качества дают в результате распределение точности оценок или скорости их получения для каждого из условий. Анализ состоит в установлении соотношения условий при основной тенденции (и дисперсии) оценок и часто использует метод анализа вариации или аналогичный метод.

Приложение 4 к части 2

Методы сравнения воздействий

В случае методов сравнения воздействий демонстрируются два изображения или две последовательности изображений, и зритель оценивает соотношение этих двух демонстраций.

A4-1 Общий порядок

Условия способа просмотра, исходные сигналы, диапазон условий и привязки, наблюдатели, ознакомление с оценкой и представление результатов определяются или выбираются в соответствии с разделом 2 части 1.

A4-2 Выбор материала для испытаний

Используемые изображения и последовательности изображений создаются тем же способом, что и для метода с одним источником воздействия. Полученные изображения и последовательности изображений затем комбинируются по парам, которые будут использованы в экспериментах по оценке.

A4-3 Сеанс испытаний

В эксперименте по оценке будет использоваться либо один, либо два дисплея с хорошим качеством, и он будет проходить, как правило, аналогично случаю с одним источником воздействия. Если используется один дисплей, то в эксперименте будет использоваться дополнительное поле воздействия такой же длительности, что и первое. В этом случае полезно убедиться на практике, что в процессе испытаний оба компонента пары появляются одинаково часто как в первой, так и во второй позиции. Если используются два дисплея, то поля воздействия демонстрируются одновременно.

Методы сравнения воздействий оценивают взаимосвязь между условиями в том случае, когда сравниваются все возможные пары условий. Однако если это требует слишком большого количества наблюдений, то может быть допустимым разделение наблюдателей между оценщиками или использование выборки из всех возможных пар.

A4-4 Типы методов со сравнением воздействий

В телевизионных оценках используется три типа методов со сравнением воздействий.

A4-4.1 Методы оценок с использованием определенных категорий

В методах оценок с использованием определенных категорий наблюдатели устанавливают связь между компонентами пары и одной категорией из предлагаемого набора, которые, как правило, определены в семантических терминах. Эти категории могут описывать наличие видимых различий (например, ОДИНАКОВЫЕ, РАЗЛИЧНЫЕ), существование и направление воспринимаемых различий (например, МЕНЬШЕ, ТО ЖЕ, БОЛЬШЕ) или суждения о продолжительности и направлении. Шкала сравнения МСЭ-R показана в таблице 2-2.

ТАБЛИЦА 2-2

Шкала сравнения

-3	Намного хуже
-2	Хуже
-1	Несколько хуже
0	То же
+1	Несколько лучше
+2	Лучше
+3	Намного лучше

Этот метод позволяет получить распределение оценок по шкале категорий для каждой пары условий. Способ анализа полученных ответов зависит от сделанных оценок (например, различие) и требуемой информации (например, только достойные внимания различия, градации условий, "дистанции" между условиями испытаний и т. д.).

A4-4.2 Методы оценок без использования категорий

В случае вынесения оценок без использования категорий наблюдатели присваивают значение соотношению между составляющими пары оценок. Существует две формы такого метода:

- При использовании непрерывной шкалы оценщик присваивает каждому соотношению цифру на линии, проведенной между двумя метками (например, ТО ЖЕ – ОТЛИЧНОЕ или между крайними категориями шкалы, показанной в таблице 4). Шкалы могут включать дополнительные эталонные метки на промежуточных точках. Расстояние от одного конца линии принимается в качестве значения для каждой пары условий.
- В случае второй формы оценщик присваивает каждому соотношению цифру, которая отражает уровень его оценки в некоторой области (например, различие в качестве). Диапазон используемых цифр может быть ограниченным или неограниченным. Цифра может описывать соотношение в "абсолютном" выражении или по отношению к "стандартной" паре.

Обе формы приводят в результате к распределению цифр для каждой пары условий. Используемый метод анализа зависит от характера суждений и требуемой информации.

A4-4.3 Методы оценки качества

В некоторых случаях количественная оценка качества может быть получена на основе процедур сравнения воздействий. В методе принудительного выбора пара готовится так, что один из сигналов содержит особый уровень некоторого атрибута (например, искажений), тогда как другой может либо

содержать, либо не содержать некоторый иной уровень данного атрибута. Наблюдателю предлагается решить, какой сигнал содержит более высокий/низкий уровень данного атрибута, либо какой сигнал содержит какой-либо атрибут; точность и скорость определения принимается в качестве оценки соотношения между компонентами одной пары.

Приложение 5 к части 2

Непрерывная оценка качества при одном источнике воздействия (SSCQE)

Внедрение методов сжатия в цифровом телевидении обусловит появление ухудшений качества изображения, которые зависят от сцены и изменяются во времени. Даже во фрагментах кодированного цифровым способом видеоизображения качество может изменяться довольно в больших пределах в зависимости от содержания сцены, а искажения могут быть очень короткими. Одни только обычные методики МСЭ-R не достаточны для оценки этого типа материала. Кроме того, метод лабораторного испытания с двойным воздействием не воспроизводит домашних условий просмотра с одним источником воздействия (SS). Поэтому считается полезной проводить непрерывную количественную оценку субъективного качества кодированных цифровым способом видеоизображений, при этом участники просматривают материал только один раз без эталонного источника.

В результате была разработана и испытана методика непрерывной оценки качества при одном источнике воздействия (SSCQE).

A5-1 Записывающее устройство и установка

Для записи непрерывной оценки качества, получаемой от участников, должно использоваться ручное электронное записывающее устройство, соединенное с компьютером. Это устройство должно иметь следующие характеристики:

- ползунковый механизм без пружинящего положения;
- линейный диапазон движения длиной 10 см;
- закрепленное или устанавливаемое на столе;
- выборки записываются два раза в секунду.

A5-2 Общая форма испытательного протокола

Участники должны присутствовать на сеансах испытаний следующего формата:

- *Фрагмент программы (ФП)*: ФП соответствует одному типу программы (например, спорт, новости, постановочная программа), обработанной в соответствии с одним из испытываемых параметров качества (ПК) (например, скорость в битах); каждый ФП должен длиться не менее 5 минут.
- *Сеанс испытаний (СИ)*: СИ – это последовательность одного или нескольких сочетаний ФП/ПК без разделения, имеющая псевдослучайный порядок. В каждом СИ хотя бы один раз имеют место все ФП и ПК, но необязательно все сочетания ФП/ПК; каждый СИ должен иметь продолжительность от 30 до 60 минут.
- *Демонстрация испытания (ДИ)*: ДИ – это испытание в полном объеме. ДИ может быть разделена на СИ, чтобы были удовлетворены требования в отношении максимальной продолжительности и с целью оценки качества по всем парам СП/ПК. Если количество пар СП/ПК ограничено, то ДИ может быть осуществлена путем повторения того же СИ для выполнения испытания в течение достаточного длительного периода времени.

Для оценки качества обслуживания может быть введен аудиосигнал. В этом случае выбор сопровождающего аудиоматериала следует осуществлять перед проведением испытания, придавая ему такое же значение, что и выбору видеоматериала.

В простейшем формате испытания будет использоваться один ФП и один ПК.

А5-3 Параметры просмотра

Условия просмотра должны быть условиями, указанными в настоящее время в части 1, или специальными условиями, указанными в части 3.

А5-4 Шкалы оценок

При ознакомлении с испытанием участники должны быть поставлены в известность, что диапазон движения ползункового механизма соответствует шкале непрерывного качества, описанной в пункте 5.4 части 1.

А5-5 Наблюдатели

Не менее 15 участников, не являющихся экспертами, должны быть привлечены при условиях, в настоящее время рекомендованных в пункте 2.5 части 1.

А5-6 Инструкции для наблюдателей

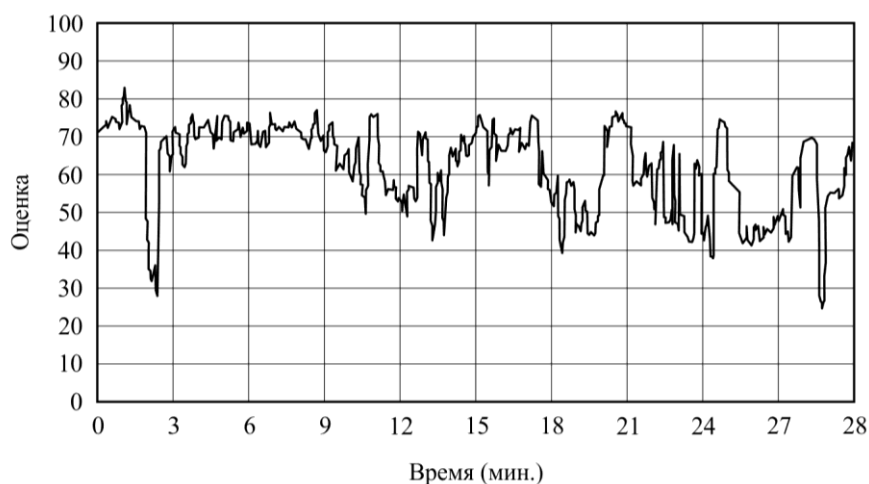
В случае оценки качества обслуживания (при звуковом сопровождении) наблюдатели должны быть проинструктированы о том, что они должны учитывать общее качество, а не только качество видеоизображения.

А5-7 Представление данных, обработка и представление результатов

Данные всех сеансов испытаний должны быть сопоставлены. Таким образом, может быть получен один график средней оценки качества в зависимости от времени $q(t)$, представляющий среднее значение всех оценок качества, выставленных наблюдателями по каждому фрагменту программы, параметру качества или всему сеансу испытаний (см., например, на рисунке 2-6).

РИСУНОК 2-6

Условие испытания. Кодек X/Фрагмент программы: Z



ВТ.0500-02-6

Однако различная задержка времени ответов разных зрителей может влиять на результаты оценки, если только рассчитывается среднее значение по фрагменту программы. Были проведены исследования для оценки воздействия времени ответа различных зрителей на результирующий класс качества.

Эти данные могут быть преобразованы в гистограмму вероятности $P(q)$ появления уровня качества q (см. пример на рисунке 2-7).

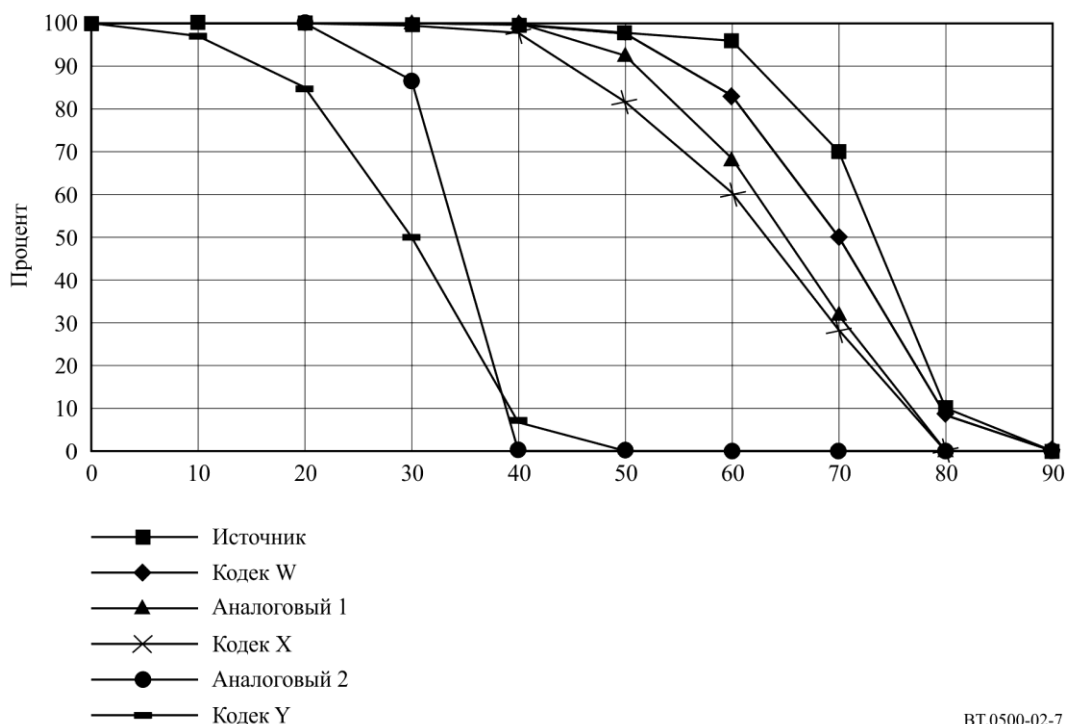
А5-8 Калибровка результатов непрерывной оценки качества и получение общей оценки качества

Тогда как было показано, что при продолжительных сеансах DSCQS получения общей оценки качества кодированного цифровым способом видеоизображения может иметь место обусловленная памятью необъективность, недавно было подтверждено, что такие явления не имеют значения при оценках DSCQS видеофрагментов продолжительностью 10 секунд. Следовательно, возможным следующим этапом в процессе SSCQE, который в настоящее время изучается, будет калибровка гистограммы качества с использованием существующего метода DSCQS на репрезентативных выборках продолжительностью 10 секунд, полученных по данным гистограммы.

Обычные методики МСЭ-R, которые использовались в прошлом, позволяли получать общие оценки качества для телевизионных последовательностей. Были проведены эксперименты, в ходе которых рассматривалась взаимосвязь между непрерывной оценкой кодированной последовательности видеоизображений и общей единой оценкой качества того же фрагмента. Уже было определено, что влияние человеческой памяти может искажать оценки качества, если значительные искажения возникают в течение 10–15 последних секунд последовательности. Однако было также обнаружено, что это влияние человеческой памяти можно смоделировать в виде убывающей экспоненциальной весовой функции. Таким образом, возможным третьим этапом методики SSCQE будет обработка этих оценок непрерывного качества с целью получения эквивалентной общей количественной оценки качества. Этот вопрос в настоящее время изучается.

РИСУНОК 2-7

Средние оценки последовательностей фрагмента программы Z, по которым проводится голосование



Приложение 6 к части 2

Метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE)

Идея проведения непрерывной оценки пришла МСЭ-R, поскольку предыдущие методы содержали некоторые несоответствия измерениям качества видеоизображений в случае схем цифрового сжатия. Основные недостатки предыдущих стандартизированных методов были связаны с возникновением контекстных артефактов на демонстрируемых цифровых изображениях. В предыдущих протоколах продолжительность времени просмотра оцениваемых видеоизображений была, как правило, ограничена 10 секундами, что, очевидно, недостаточно для наблюдателя, чтобы дать репрезентативную оценку тому, что может произойти при реальном обслуживании. Цифровые артефакты сильно зависят от пространственного и временного содержания изображения источника. Это является верным для схем сжатия, но также касается поведения цифровых систем передачи в отношении устранения ошибок. В случае прошлых стандартизированных методик было очень сложно выбирать последовательности типовых видеоизображений или, по крайней мере, оценивать их репрезентативность. По этой причине МСЭ-R внедрил метод SSCQE, с помощью которого возможно измерять качество видеоизображений на длинных последовательностях, которые являются репрезентативными по содержанию видеоизображений и статистическим данным ошибок. В целях воспроизведения условий просмотра, как можно более приближенных к реальным ситуациям, в методе SSCQE не используются эталонные изображения.

При необходимости оценить верность следует ввести эталонные изображения. Метод SDSCE был разработан на основе метода SSCQE путем введения небольших отступлений, касающихся способа демонстрации изображений участникам и шкалы оценок. Метод был предложен группе MPEG для оценки устойчивости к ошибкам при очень низких скоростях передачи в битах, однако он может быть подходящим для применения во всех случаях, когда следует оценить верность зрительной информации, затронутой изменяемым во времени ухудшением.

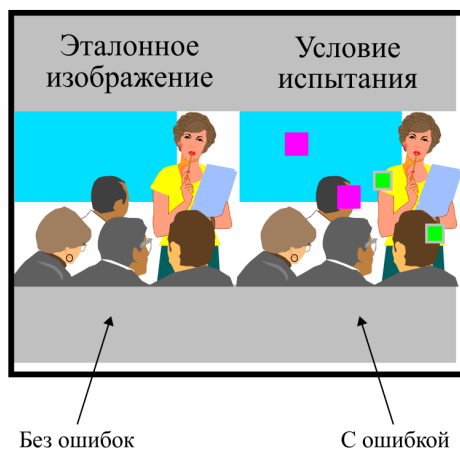
В результате был разработан и испытан следующий новый метод SDSCE.

А6-1 Процедура испытания

Группа участников одновременно просматривает две последовательности: одну – эталонную, другую – относящуюся к одному из условий испытания. Если формат последовательностей SIF (формат стандартного изображения) или формат меньшего размера, то обе последовательности могут отображаться рядом на одном дисплее, в ином случае должны использоваться два рядом стоящих дисплея (см. рисунок 2-8).

РИСУНОК 2-8

Пример формата устройства отображения



ВТ.0500-02-8

Участникам предлагается выявить различия между двумя последовательностями и оценить верность видеоинформации путем перемещения ползунка ручного устройства голосования. Если верность безусловная, то ползунок должен быть вверху шкалы оценок (кодировка 100), а если верность нулевая, то внизу шкалы (кодировка 0).

Участники проинформированы, на каком экране демонстрируется эталонное изображение, и им предлагается выразить свое мнение в процессе просмотра последовательностей на всем протяжении их демонстрации.

А6-2 Различные этапы

Этап обучения – это важнейшая часть данного метода испытаний, поскольку участники могут неправильно иметь свою задачу. Должны быть предоставлены инструкции в письменном виде для обеспечения уверенности в том, что все участники получают абсолютно одинаковую информацию. Инструкции должны включать пояснение о том, что участники увидят, что они должны оценивать (т. е. разницу в качестве) и как они выражают свое мнение. На все вопросы участников должны быть даны ответы в целях недопущения, по мере возможности, любого недопонимания мнения руководителем испытания.

После предоставления инструкций должен быть проведен *демонстрационный сеанс*. Таким способом осуществляется ознакомление участников с процедурами голосования и видом искажений.

Наконец, должно быть проведено моделирование испытания, при котором демонстрируется ряд типичных условий. Последовательности должны отличаться от используемых в испытании и должны воспроизводиться одна за другой без перерыва.

После того как *моделирование испытания* завершено, экспериментатор должен, главным образом, проверить, что в случае, когда условия испытаний эквивалентны эталонным условиям, оценки близки к "сотне" (то есть разница не была замечена); если же вместо этого участники заявляют о том, что видят некоторые различия, экспериментатор должен повторить объяснение и моделирование испытания.

А6-3 Особенности протокола испытаний

Следующие определения применяются к описанию протокола испытаний:

- *Фрагмент видеозображения (ФВ)*: ФВ соответствует одной последовательности видеозображений.
- *Условие испытания (УИ)*: УИ может быть конкретный процесс, связанный с видеозображением, условие передачи или они оба. Каждый ФВ должен обрабатываться в соответствии, по крайней мере, с одним УИ. Кроме того, в список УИ должны быть добавлены эталоны для образования оцениваемых пар эталон/эталон.

- *Сеанс (С)*: сеанс – это последовательность различных пар СВ/УИ без разделения, построенная в псевдослучайном порядке. В каждом сеансе хотя бы один раз встречаются все ФВ и УИ, но необязательно все сочетания ФВ/УИ.
- *Испытательная демонстрация (ИД)*: испытательная демонстрация – это последовательность испытаний, охватывающих все сочетания ФВ/УИ. То же число наблюдателей (но не обязательно те же наблюдатели) должно голосовать по всем сочетаниям ФВ/УИ.
- *Период вынесения решений*: в ходе сеанса каждому наблюдателю предлагается непрерывно выносить решения.
- *Сегмент вынесения решений (СПР)*: 10-секундный сегмент вынесения решений; все СПР получают путем использования групп 20 последовательных решений (эквивалентных 10 секундам) без какого-либо перекрытия.

А6-4 Обработка данных

После проведения испытания будет получен один (или несколько) файл (файлов), содержащий (содержащие) все решения, которые получены в ходе различных сеансов (С), представляющих общее количество решений в отношении ИД. Первое выяснение действительности данных может быть выполнено путем проверки того, что каждая пара ФВ/УИ была рассмотрена и что равное количество решений было вынесено по каждой из них.

Данные, собранные в ходе испытаний, которые проведены в соответствии с этим протоколом, могут быть обработаны тремя различными способами:

- статистический анализ каждого отдельного ФВ;
- статистический анализ каждого отдельного УИ;
- общий статистический анализ всех пар ФВ/УИ.

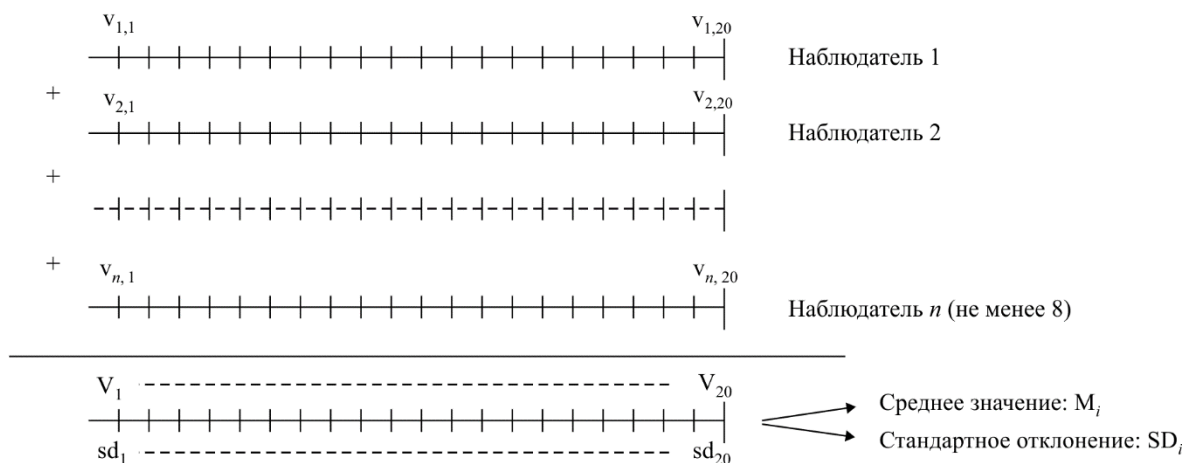
В каждом случае необходим многошаговый анализ:

- Средние значения и стандартные отклонения рассчитываются по каждому решению путем накопления наблюдателей.
- Средние значения и стандартное отклонение рассчитываются для каждого СПР, как показано на рисунке 2-9. Результаты этого этапа могут быть представлены на временной диаграмме, как показано на рисунке 2-10.
- Анализируется статистическое распределение средних значений, рассчитанных на предыдущем этапе (то есть соответствующем каждому СПР), и частота их появления. В целях недопущения эффекта новизны, обусловленного предыдущими сочетаниями ФВ × УИ, первые десять СПР для каждой выборки ФВ × УИ отбрасываются.
- Общая характеристика раздражающего воздействия рассчитывается путем накопления данных о частоте появления. В этом расчете должны быть учтены доверительные интервалы, как показано на рисунке 2-11. Общая характеристика раздражающего воздействия соответствует этой статистической интегральной функции распределения, указывая соотношение между средними значениями для каждого сегмента вынесения решения и их кумулятивной частотой появления.

РИСУНОК 2-9

Обработка данных

а) Расчет средней оценки V и стандартного отклонения SD в момент вынесения решения наблюдателем для каждой последовательности, по которой проводится голосование, каждого сочетания ФВ \times УИ



б) Расчет M и SD для каждой последовательности, по которой проводится голосование, продолжительностью 1 секунда для каждого сочетания ФВ \times УИ



ВТ.0500-02-9

А6-5 Надежность участников

Надежность участников может быть качественно оценена путем выяснения их поведения при демонстрации пар эталон/эталон. В этих случаях предполагается, что участники дадут оценки, весьма близкие к 100. Это доказывает, что они, по крайней мере, понимают свою задачу и не выносят случайных решений.

Кроме того, надежность участников может быть проверена путем использования процедур, близких к методу SSCQE, описанному в пункте А1-2.3.2 Приложения 1 к части 1.

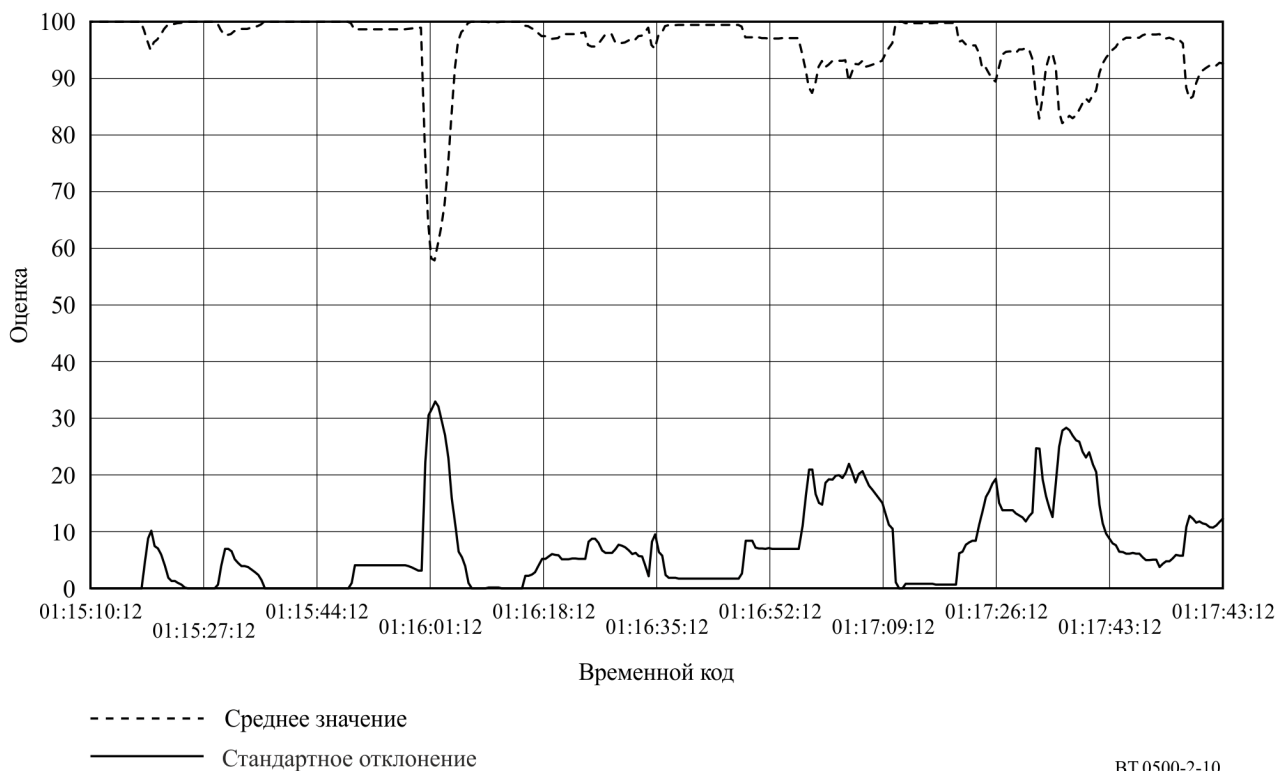
В случае процедуры SDSCE надежность решений зависит от следующих двух параметров:

Систематические отклонения: в ходе испытания зритель может быть слишком оптимистичным или слишком пессимистичным или может даже неправильно понимать процедуры вынесения решений (например, шкалу оценок). Это может привести к более или менее систематическому отклонению последовательности решений от средней последовательности, если не полному выходу за пределы диапазона.

Местные отклонения от нормального порядка: как и в других хорошо известных испытательных процедурах наблюдатели иногда могут выносить решения, не слишком внимательно просматривая демонстрируемые последовательности и не слишком тщательно отслеживая их качество. В этом случае общая кривая решений может относительно находиться в пределах усредненного диапазона. Однако, несмотря на это, могут наблюдаться местные отклонения от нормального порядка.

Эти два нежелательных эффекта (нетипичное поведение и отклонения от нормального порядка) можно избежать. Разумеется, весьма важным является обучение участников. Однако следует сделать возможным использование средства, позволяющего выявлять несоответствующих наблюдателей и, при необходимости, отстранять их. Предложение в отношении двухступенчатого процесса, позволяющего осуществлять такую фильтрацию, описано в настоящей Рекомендации.

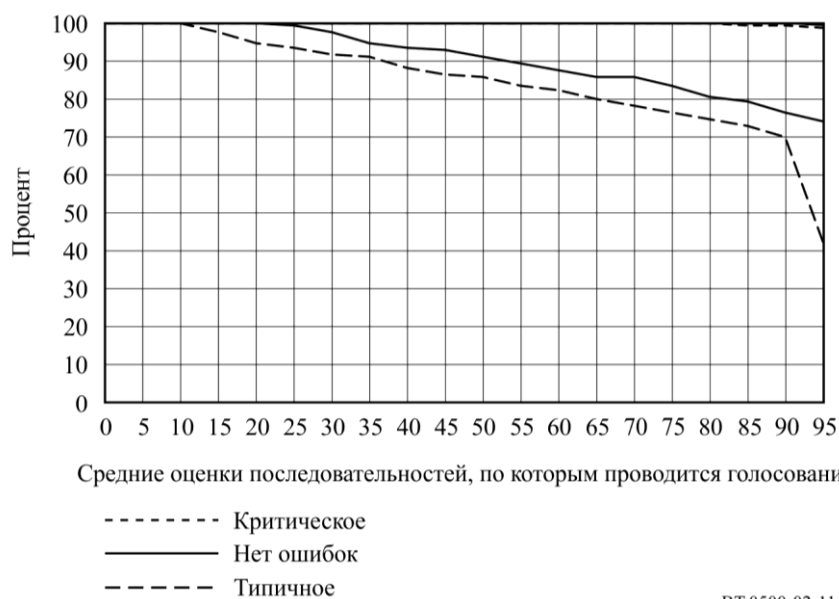
РИСУНОК 2-10
Групповая временная диаграмма



ВТ.0500-2-10

РИСУНОК 2-11

Общие характеристики раздражающего воздействия, рассчитанные на основе статистических распределений и включающие доверительный интервал



ВТ.0500-02-11

Приложение 7 к части 2

Субъективная оценка качества мультимедийного видеоизображения (SAMVIQ)

A7-1 Введение

В методе оценки качества SAMVIQ используется шкала непрерывного качества для обеспечения измерения подлинного качества последовательностей видеоизображений. Каждый наблюдатель передвигает скользящий маркер по непрерывной шкале, отградуированной от 0 до 100, на которой отмечены пять линейно расположенных оценок качества (отлично, хорошо, удовлетворительно, плохо, неприемлемо).

В методе SAMVIQ зрителю предоставляется возможность получить доступ к нескольким вариантам последовательности. После того как зрителем дана оценка всех вариантов, может быть предоставлен доступ к содержанию следующей последовательности.

Различные варианты выбираются зрителем случайным образом с помощью графического компьютерного интерфейса. Зритель может сколько угодно раз останавливать, просматривать заново или изменять оценку каждого варианта последовательности. Данный метод включает явную эталонную (то есть необработанную) последовательность, а также несколько вариантов той же самой последовательности, которые содержат как обработанные, так и не обработанные последовательности (то есть скрытый эталон). Каждый вариант последовательности демонстрируется отдельно и оценивается с использованием шкалы непрерывного качества, аналогичной используемой в методе DSCQS. Следовательно, этот метод функционально весьма близок к методу с одним входным сигналом и со случайным доступом, однако наблюдатель может просматривать явный эталон, когда он захочет, что делает данный метод аналогичным тому, в котором используется эталонная последовательность.

Оценка качества выполняется *цена за ценой* (см. рисунок 2-12), включая *явный эталон*, *скрытый эталон* и *различные алгоритмы*.

Для лучшего понимания этого метода ниже определяются следующие конкретные слова:

Цена: аудиовизуальный контент;

Последовательность: сцена, смешанная с обработанными частями, или без них;

Алгоритм: один или несколько методов обработки изображений.

A7-2 Явный, скрытый эталоны и алгоритмы

Метод оценки обычно включает привязки качества для обеспечения устойчивости результатов. По изложенным ниже причинам в методе SAMVIQ учитываются две привязки, относящиеся к высокому качеству. Было проведено несколько испытаний, которые продемонстрировали, что стандартное отклонение оценок уменьшается при использовании *явного эталона*, а не при скрытом эталоне или при отсутствии эталона. В частности, для получения максимальной надежности результатов при оценке показателей работы кодека лучше использовать явный эталон. Для оценки подлинного качества вместо явного добавляется также *скрытый эталон*, поскольку демонстрация и обработанные последовательности являются анонимными. Явное название "эталона" оказывает влияние примерно на 30% наблюдателей. Эти наблюдатели дают наивысшую оценку (100) явному эталону, которая полностью отличается от соответствующей оценки скрытого эталона. В частности, при отсутствии в распоряжении эталона проведение испытания остается возможным, однако стандартное отклонение существенно возрастает.

Использование метода SAMVIQ целесообразно для мультимедийного содержания, поскольку можно объединять различные особенности обработки изображения, например тип кодека, формат изображения, скорость передачи данных в битах, временное обновление, наплывы изображения и т. п. Одна из таких особенностей или их сочетание кратко называются *алгоритмом*.

А7-3 Условия проведения испытаний

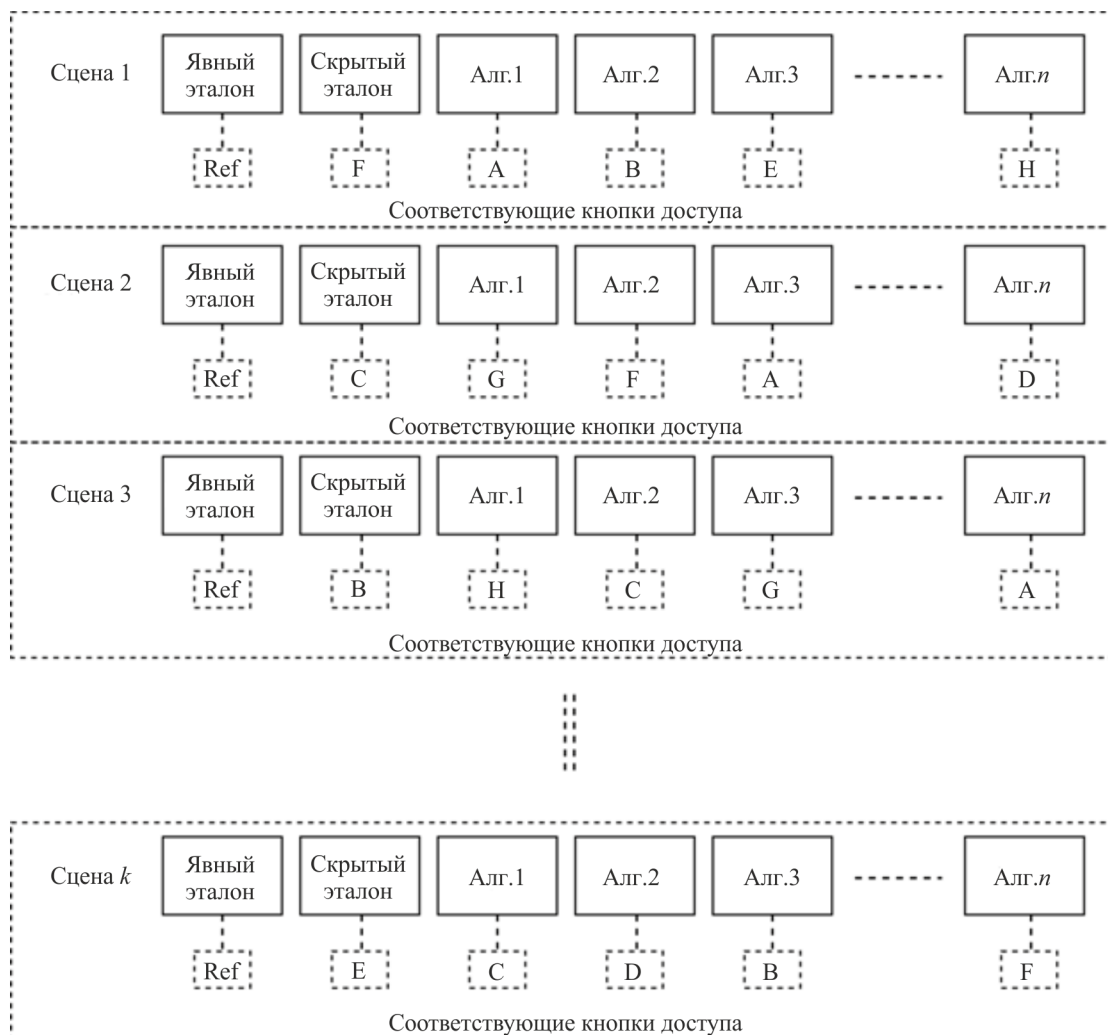
Изменение критичности на протяжении сцены ограничено, поскольку содержание выбирается однородным в соответствии с теми же правилами, которые неявно используются в других методах, обеспечивающих общую оценку (например, в методах с одним входным сигналом). Поэтому максимальная длительность просмотра последовательности, составляющая 10 или 15 секунд, является достаточной для получения устойчивой и надежной оценки качества. Для сохранения надлежащего качества отображения следует использовать фирменные декодеры-устройства воспроизведения или экранную копию их выходного сигнала.

А7-4 Организация испытаний

- a) Испытание проводится по сценам, как показано на рисунке 2-12.
- b) Для текущей сцены можно воспроизводить и оценивать любую последовательность в любом порядке. Каждая последовательность может воспроизводиться и оцениваться несколько раз.
- c) При переходе от одной сцены к другой доступ к последовательностям осуществляется случайным образом и предотвращает попытки наблюдателей голосовать одинаковым образом в соответствии с установленным порядком. В действительности порядок алгоритма для данного испытания остается одним и тем же для упрощения анализа и представления результатов. Случайный порядок устанавливается только для соответствующего доступа с помощью идентичной кнопки.
- d) При первом просмотре текущая последовательность обязательно должна быть полностью воспроизведена до выставления оценки, в ином случае можно было бы немедленно дать оценку и остановиться.
- e) Для проведения испытания следующей сцены должна быть выставлена оценка всем последовательностям текущей сцены.
- f) Для завершения испытания обязательно должны быть оценены все последовательности всех сцен.

РИСУНОК 2-12

Пример организации испытания для метода SAMVIQ



ВТ.0500-02-12

Метод SAMVIQ реализуется с помощью программного обеспечения. Кроме кнопок доступа, показанных на рисунке 2-12, необходимы кнопки "воспроизведение", "стоп", "следующая сцена" и "предыдущая сцена" для предоставления зрителю возможности управлять демонстрацией различных сцен (см., например, пункт А7-6). Оценка, выставленная зрителем, должна быть показана под кнопкой доступа, соответствующей этой сцене. При необходимости после выставления оценок всем различным вариантам последовательности зрителю все еще предоставляется возможность сравнивать и изменять значения оценок. Нет необходимости анализировать всю текущую последовательность, поскольку существенные различия уже были выделены в ходе первого состоявшегося просмотра.

А7-5 Демонстрация и анализ данных

А7-5.1 Итоговая информация

Точная информация об условиях испытаний необходима для повторения того или иного испытания или для сравнения результатов различных испытаний. Поэтому рекомендуется записывать информацию об условиях проведения испытания, представленную в таблице 2-3.

ТАБЛИЦА 2-3

Итоговая информация об испытании

Название метода	
Технология устройства отображения	
Название устройства отображения для справки	
Пиковый уровень яркости (кд/м ²)	
Уровень яркости черного (кд/м ²)	
Установка уровня черного: испытательный сигнал PLUGE (порог восприятия расстояния между уровнем черного и уровнем "чернее черного" = 8). В ином случае указывается значение порога.	
Уровень яркости фона (кд/м ²)	
Освещенность (люкс)	
Расстояние просмотра: – не ограниченное: передняя сторона устройства отображения – ограниченное: nH	
Размер устройства отображения (по диагонали в дюймах)	
Отношение ширины к высоте устройства отображения	
Формат устройства отображения (число колонок и строк)	
Формат входного изображения (число колонок и строк)	
Формат выходного изображения ⁽¹⁾ (число колонок и строк)	
Цветовая температура белого: D65, в ином случае Координаты цвета белого (x, y)	
Число действующих наблюдателей	

⁽¹⁾ Данная информация требуется, когда входное изображение обрабатывается при его отображении, например изменяется его масштаб.

Характеристики устройства отображения могут оказывать влияние на результаты испытаний. Для плоских экранов необходима дополнительная информация, например о характеристике яркости (гамма-точность) и основных цветах.

Характеристики последовательностей видеоизображения необходимы для планирования эксперимента или объяснения его результатов. Рекомендуется предоставлять данные о пространственно-временных характеристиках, как описано в Приложении 1 к части 1. Эта информация должна приниматься во внимание при сборе испытательных последовательностей в библиотеку видеоматериалов, подходящих для субъективной оценки качества видеоизображения в мультимедийных приложениях.

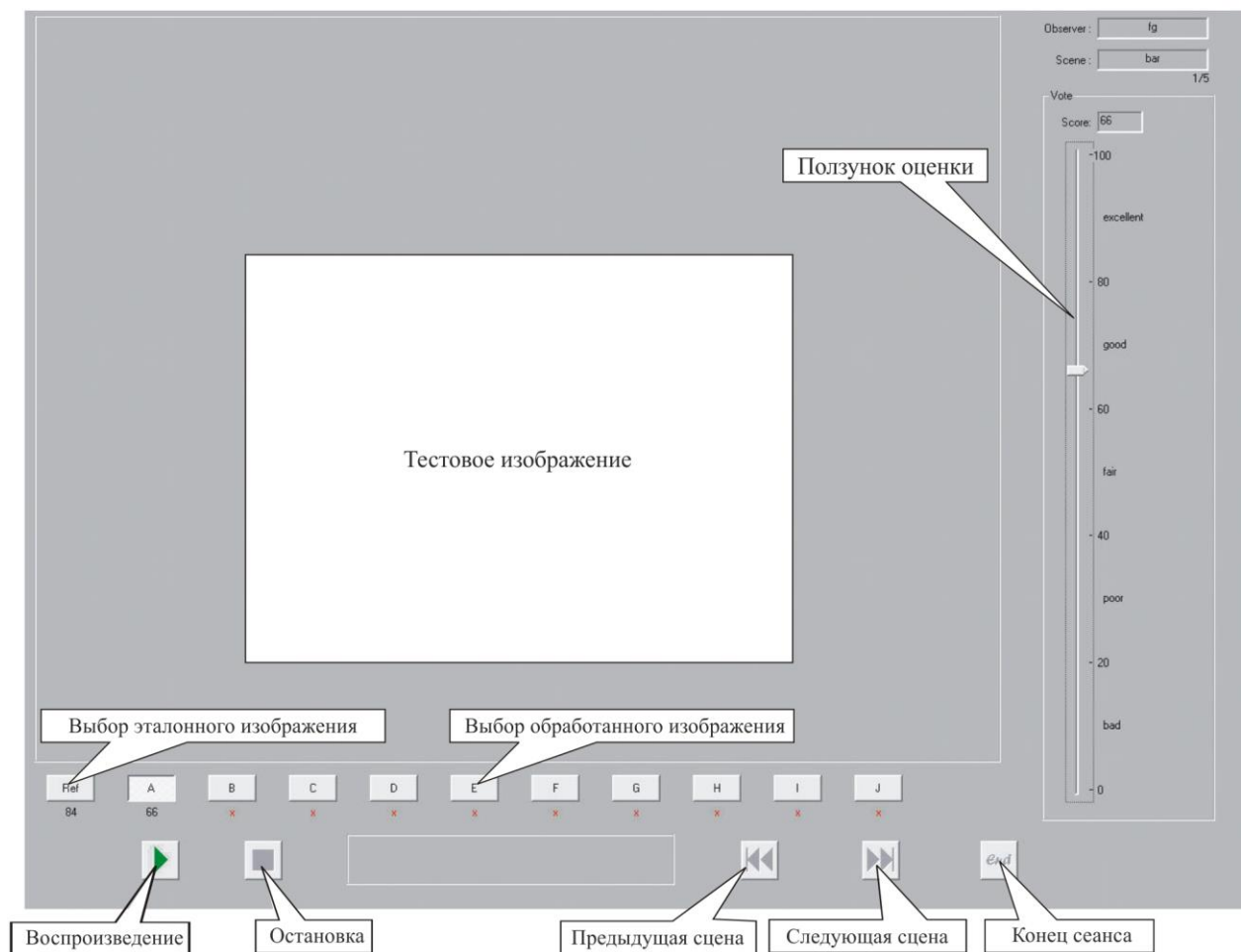
A7-5.2 Методы анализа

Методами анализа являются методы, описанные в Приложении 1 к части 1.

A7-5.3 Отбор наблюдателей

Отбор для SAMVIQ описан в разделе A1-2 Приложения 1 к части 1.

А7-6 Пример интерфейса SAMVIQ (информационный)



ВТ.0500-02-12a

Приложение 8 к части 2

Протокол просмотра экспертами (EVP) для оценки качества видеоматериала

В настоящем Приложении описан метод субъективной оценки видеокачества движущихся изображений с помощью протокола просмотра экспертами при участии ограниченного числа зрителей, которые все отобраны из числа экспертов в соответствующей области обработки видео.

А8-1 Организация лабораторных условий

А8-1.1 Выбор дисплея и настройка

В качестве дисплея следует использовать дисплей с плоским экраном, характеристики которого аналогичны типовым характеристикам профессиональных приложений (например, телевизионная студия или передвижная телевизионная станция); размер дисплея по диагонали может варьироваться от 22 дюймов (минимальное значение) до 40 дюймов (рекомендуемое значение), но может также составлять до 50 дюймов и выше в случае оценки систем с разрешающей способностью ТВЧ или выше.

Разрешается использовать уменьшенную часть активной области просмотра дисплея, но в этом случае цвет области вокруг активной части дисплея следует установить средне-серым. При таком условии использования следует запрещать установку разрешающей способности дисплея, отличную от его собственной.

Дисплей должен обеспечивать надлежащую настройку и калибровку яркости и цвета с использованием профессионального люксметра. Калибровка дисплея должна соответствовать параметрам, определенным в соответствующей Рекомендации для проводимого испытания.

A8-1.2 Расстояние просмотра

Расстояние просмотра, на котором располагаются эксперты, следует выбирать в соответствии с разрешающей способностью экрана и высотой активной части экрана, в соответствии с проектным расстоянием просмотра, которое описано в пункте 2.1.3.2 части 1, или выбирать меньшее расстояние просмотра в соответствии с требованиями, определяемыми критическими условиями просмотра.

A8-1.3 Условия просмотра

Эксперимент для получения протокола просмотра экспертами (EVP) необязательно проводить в испытательной лаборатории, но важно, чтобы место проведения испытаний было защищено от видимых и слышимых помех (например, можно использовать тихое служебное помещение или комнату для переговоров).

Следует устранить любой источник прямого или отраженного света, падающего на экран; другой окружающий свет должен быть тусклым, поддерживаемым на минимальном уровне, позволяющем заполнять оценочные листы (если таковые используются).

Число сидящих перед дисплеем экспертов может варьироваться в зависимости от размера экрана, с тем чтобы гарантировать одинаковое для зрителей воспроизведение изображения и представление входных последовательностей.

A8-2 Зрители

Зрители, участвующие в эксперименте для получения EVP, должны быть экспертами в предметной области исследования.

Необязательно проверять остроту зрения и дальтонизм зрителей, так как они должны быть отобраны из числа квалифицированных лиц.

Минимальное число разных зрителей должно составлять девять.

Для охвата минимального числа зрителей тот же эксперимент может быть проведен в том же месте с повторением испытаний или в нескольких местах. Оценки, полученные в разных местах проведения одного сеанса просмотра экспертами, могут быть статистически обработаны вместе.

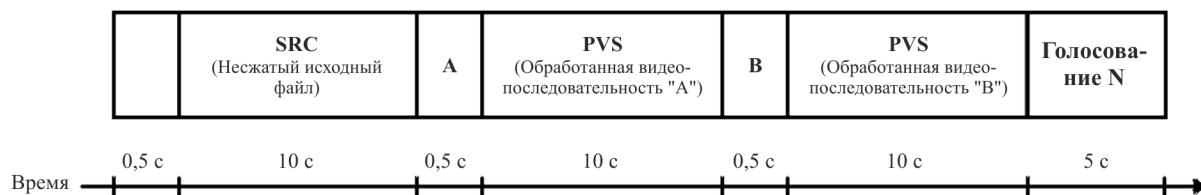
A8-3 Базовая ячейка испытаний

Представляемый экспертам материал должен быть организован с созданием базовой ячейки испытаний (BTC) для каждой пары оцениваемых условий кодирования (см. рисунок 2-13).

Рассматриваемые в рамках BTC фрагменты исходных эталонных последовательностей (SRC) и фрагменты обработанных видеопоследовательностей (PVS) всегда должны относиться к той же видеопоследовательности, с тем чтобы эксперты могли определить любое улучшение качества изображения, обеспечиваемое испытываемыми алгоритмами сжатия.

РИСУНОК 2-13

Временные интервалы базовой ячейки испытаний для протокола просмотра экспертами



ВТ.0500-02-13

ВТС должна быть организована следующим образом:

- 0,5 с – экран, цвет которого установлен средне-серым (среднее значение шкалы яркости);
- 10 с – представление эталонного несжатого видеофрагмента;
- 0,5 с – показ сообщения "А" (первое видео для оценки) на средне-сером фоне;
- 10 с – представление ухудшенной версии видеофрагмента;
- 0,5 с – показ сообщения "В" (второе видео для оценки) на средне-сером фоне;
- 10 с – представление ухудшенной версии видеофрагмента;
- 5 с – показ сообщения, предлагающего зрителям выразить свое мнение.

Сообщение "Голосование" должно сопровождаться числом, которое помогает ориентироваться в оценочном листе.

A8-4 Оценочный лист и шкала оценок

Как показано на рисунке 2-13, представление видеофрагментов должно быть организовано таким образом, чтобы первой была показана неухудшенная эталонная последовательность (SRC), а далее – две ухудшенные видеопоследовательности (PVS). Порядок представления PVS должен произвольно меняться для каждой ВТС, и зрители не должны знать порядок представления.

РИСУНОК 2-14

Пример оценочного листа для состоящего из 24 ВТС сеанса просмотра экспертами

Сеанс №										
Решение 1		Решение 2		Решение 3		Решение 4		Решение 5		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Решение 6		Решение 7		Решение 8		Решение 9		Решение 10		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Решение 11		Решение 12		Решение 13		Решение 14		Решение 15		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Решение 16		Решение 17		Решение 18		Решение 19		Решение 20		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Решение 21		Решение 22		Решение 23		Решение 24				
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
Место					Участок					
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>			

ВТ.0500-02-14

Используется 11-балльная числовая шкала от 10 (неразличимые ухудшения) до 0 (сильно раздражающие ухудшения).

В таблице 2-4 показаны значения 11-балльной числовой шкалы.

ТАБЛИЦА 2-4

Значения 11-балльной числовой шкалы

Оценка	Ухудшение	
10	Неразличимое	
9	Малоразличимое	кое-где
8		езде
7	Различимое	кое-где
6		езде
5	Очевидно различимое	кое-где
4		езде
3	Раздражающее	кое-где
2		езде
1	Чрезвычайно раздражающее	кое-где
0		езде

Зрителям предлагается заполнить формуляр, в котором для каждой ВТС предусмотрены две ячейки (помеченные как "А" и "В"), проставив в каждой из этих двух ячеек оценку, выбранную по 11-балльной числовой шкале.

На рисунке 2-14 представлен пример оценочного листа для сеанса, состоящего из 24 ВТС.

Для каждой ВТС зрители заполняют ячейку, обозначенную буквой **A** (для оценки видеофрагмента, показанного первым), и ячейку, обозначенную буквой **B** (для оценки видеофрагмента, показанного вторым).

Представление исходного не ухудшенного видеофрагмента упрощает оценку экспертами любых ухудшений.

Значения 11-балльной числовой шкалы следует подробно объяснить в ходе "тренировочных сеансов", описанных ниже.

A8-5 План испытания и формирование сеанса

Порядок представления ВТС должен устанавливаться разработчиком испытания произвольным образом, так чтобы исключить представление два раза подряд того же видеофрагмента, как и того же ухудшенного фрагмента.

Любой сеанс просмотра следует начинать с "этапа стабилизации", в который включается "лучшая" ВТС, "худшая" ВТС и две ВТС "среднего качества" наряду с прочими ВТС, включенными в каждый сеанс испытаний. Это позволит зрителям сразу получить непосредственное представление о диапазоне качества уже в начале сеанса испытаний.

Если продолжительность сеанса испытаний превышает 20 минут, разработчик испытания должен разделить его на два (или более) отдельных сеанса просмотра и каждый из них не должен превышать 20 минут. В таком случае каждый сеанс просмотра следует начинать с "этапа стабилизации".

A8-6 Подготовка участников

Даже если проведение данной процедуры запланировано с участием экспертов, перед каждым экспериментом предпочтительно проводить короткий (5–6 ВТС) тренировочный сеанс просмотра.

Для тренировочного сеанса может быть использован тот же видеоматериал, который будет использоваться в реальных сеансах, но порядок представления должен быть иным.

Зрители должны быть подготовлены к использованию 11-балльной шкалы, для чего им следует предложить внимательно просмотреть видеофрагменты, показываемые сразу после сообщений "A" и "B" на экране, и проверить, могут ли они заметить разницу по сравнению с видеофрагментом, показанным первым (SRC).

A8-7 Сбор и обработка данных

В конце каждого сеанса следует производить сбор оценок и их регистрацию в электронной таблице для расчета средних значений.

Желательно выполнить "последующее отсеивание" зрителей с использованием линейного коэффициента корреляции Пирсона.

Ко всем оценкам каждого участника следует применять функцию "корреляция" относительно средней экспертной оценки (MOS); может быть установлено пороговое значение для определения каждого зрителя как "допустимый" и "отклоненный" (в Рекомендации МСЭ-Т Р.913 предлагается использовать пороговое значение "отклоненный", равное 0,75).

A8-8 Условия использования результатов протокола просмотра экспертами

Протокол просмотра экспертами (EVP) может использоваться в тех случаях, когда время и ресурсы не позволяют проводить эксперимент для формальной субъективной оценки.

Для получения EVP требуется меньше времени, чем для проведения формальной субъективной оценки, и такой просмотр может проходить в "неформальной" среде при отсутствии видимых и слышимых внешних помех.

К единственным обязательным условиям относятся окружающее освещение и условия просмотра (дисплей, угол наблюдения и расстояние просмотра), которые описаны в предыдущих пунктах.

A8-9 Ограничения использования результатов EVP

Даже при том что EVP демонстрирует возможность обеспечения приемлемых результатов с участием всего девяти зрителей, MOS по результатам эксперимента для получения EVP не может рассматриваться в качестве замены результатов, которые возможно получить по итогам эксперимента для формальной субъективной оценки.

Данные MOS, получаемые с использованием EVP, могут использоваться как предварительное представление уровня ухудшения.

Данные MOS, получаемые с использованием EVP, могут использоваться для предварительного ранжирования оцениваемых схем обработки видео.

Если удобно или необходимо, эксперимент для получения EVP может проводиться параллельно в большем числе мест, при условии идентичности условий просмотра, расстояния просмотра и плана испытаний.

Если число экспертов-зрителей, участвующих в том же эксперименте для получения EVP, а также при проведении эксперимента в разных местах, составляет 15 или более, необработанные субъективные данные могут обрабатываться для получения MOS, среднеквадратического отклонения и доверительного интервала, что может помочь провести более точное ранжирование испытуемых элементов. В этом последнем случае может быть выполнен более точный инферентный статистический анализ, например Т-критерий Стьюдента.

Прилагаемый документ 1 к Приложению 8 (информационный) к части 2

Применение протокола на основе просмотра экспертами и его динамика при участии большого числа экспертов-оценщиков

В настоящем информационном Прилагаемом документе представлены сведения о результатах двух разных сеансов субъективной оценки кодированных видеофрагментов высокой четкости (ВЧ) и сверхвысокой четкости (СВЧ) для получения EVP, которые проводились в ходе 117-го собрания MPEG с применением положений Приложения 8, с тем чтобы быстро и надежно оценить два разных метода кодирования источника.

Ввиду присутствия большого числа экспертов, участвующих в 117-м собрании MPEG, количество оценщиков, участвовавших в двух сеансах для получения EVP, было значительно больше девяти – рекомендованного значения в Приложении 8 к части 2 настоящей Рекомендации; в испытательном сеансе оценки ВЧ для получения EVP приняли участие 30 экспертов, а в сеансе оценки СВЧ для получения EVP – 32 эксперта.

Широкое участие экспертов-оценщиков предоставило возможность проанализировать данные MOS с целью проверки уровня надежности, обеспечиваемого при использовании Приложения 8 для оценки кодированных видеофрагментов.

В оценке участвовали четыре группы зрителей (9, 12, 15 и 18 человек), и проводилось сравнение значений MOS, полученных от девяти экспертов, со значениями MOS, полученными от 12, 15 и 18 зрителей.

Цель состояла в том, чтобы сравнить оценки, полученные от девяти экспертов (то есть в соответствии с протоколом EVP), с оценками, полученными от 12, 15 и 18 экспертов (то есть аналогично эксперименту для формальной субъективной оценки).

Как видно из рисунков 2-15 (эксперимент с видеоматериалом СВЧ) и 2-16 (эксперимент с видеоматериалом ВЧ), результаты оценки во всех четырех рассмотренных случаях весьма близки.

Приняв результаты, полученные от 18 зрителей, за своего рода "реальную ситуацию", можно построить графики, представленные на рисунках 2-15 и 2-16, на которые нанесены контрольные точки в соответствии с полученными значениями MOS от 18 зрителей (сплошная красная линия).

Другие линии на графиках отражают результаты, полученные от девяти зрителей (пунктирная красная линия), 12 зрителей (синяя пунктирная линия) и 15 зрителей (сплошная зеленая линия).

Анализируя результаты, представленные на рисунках 2-15 и 2-16, можно отметить:

- графики результатов 15 и 18 зрителей отражают монотонное снижение от значений MOS высокого качества к значениям низкого качества;
- графики результатов 9 и 12 зрителей отражают некоторые "инверсии" оценок по сравнению с графиком результатов 18 зрителей, хотя эти колебания оценок довольно ограничены по величине.

В итоге, описанные здесь эксперименты для получения EVP показывают очень хорошую характеристику протокола EVP, подтверждая то, что заявлено в тексте Приложения 8, а именно что протокол EVP, хотя и не может рассматриваться в качестве полной замены формального субъективного эксперимента, он может рассматриваться в качестве стабильной процедуры оценки, обеспечивающей результаты, весьма близкие к тем, которые получены с участием гораздо большего числа зрителей при проведении формальной субъективной оценки.

РИСУНОК 2-15

Оценка, полученная в ходе эксперимента по оценке фрагментов СВЧ, в зависимости от числа оценщиков

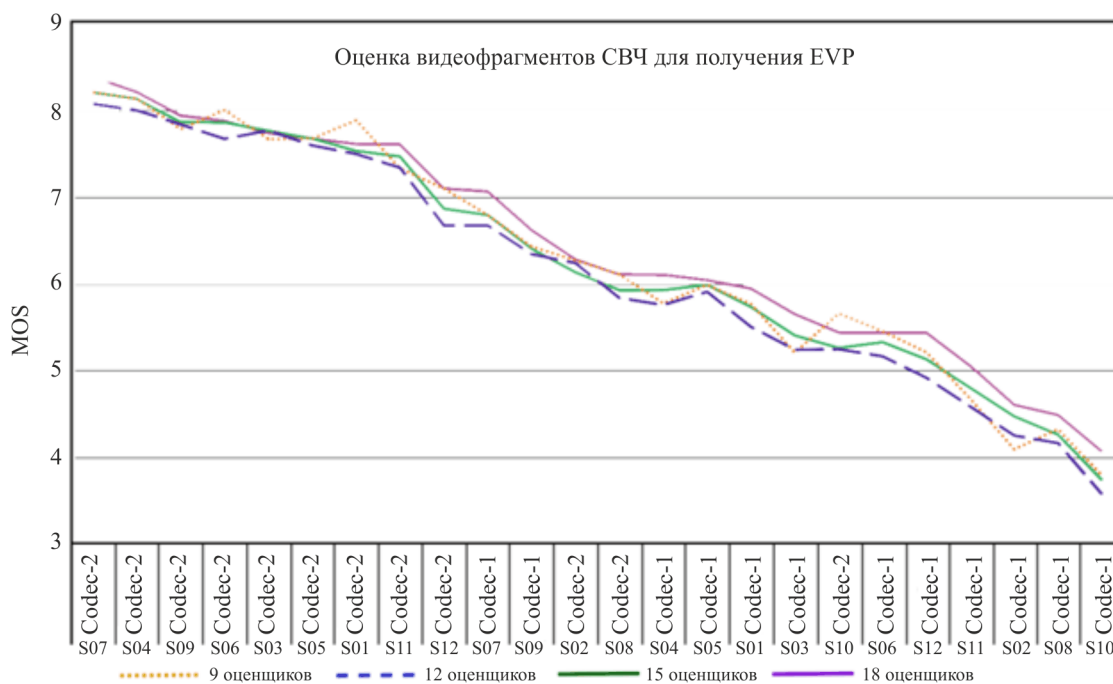
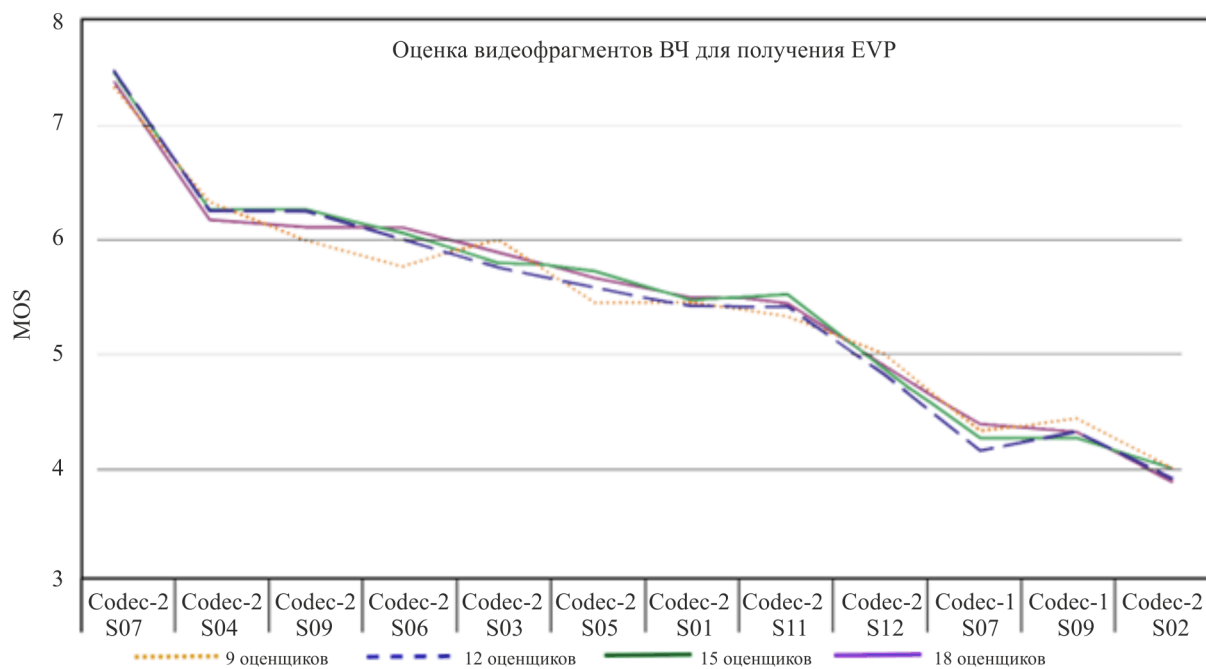


РИСУНОК 2-16

Оценка, полученная в ходе эксперимента по оценке фрагментов ВЧ, в зависимости от числа оценщиков



ЧАСТЬ 3

Специализированные методики субъективной оценки качества изображения

При проектировании испытаний для субъективной оценки качества изображения следует рассмотреть специализированные методики для определенных применений. В настоящей части 3 содержится руководство по субъективной оценке качества изображения в соответствующих форматах изображения и применениях.

Приложение 1	Субъективная оценка цифровых телевизионных систем стандартной четкости (ТСЧ)
Приложение 2	Субъективная оценка качества изображения телевизионных систем высокой четкости (ТВЧ)
Приложение 3	Субъективная оценка качества буквенно-цифровых и графических изображений в телетексте и аналогичных текстовых службах
Приложение 4	Субъективная оценка качества изображения в многопрограммных службах
Приложение 5	Экспертная оценка качества изображения систем цифрового воспроизведения видеоизображения для большого экрана в кинотеатрах
Приложение 6	Субъективная оценка качества изображения в мультимедийных применениях
Приложение 7	Методы субъективной оценки систем стереоскопического 3D-телевидения

Приложение 1 к части 3

Субъективная оценка цифровых телевизионных систем стандартной четкости (ТСЧ)

A1-1 Введение

В этом Приложении, предназначенном для использования в сочетании с частями 1 и 2 настоящей Рекомендации, подробно описаны способы применения общих методов, приведенных в Рекомендации, для субъективной оценки цифровых систем, обеспечивающих уровни качества обычных телевизионных систем или близкие к ним. Приведенные здесь подробные описания процедур вместе с соответствующей базовой информацией относятся к испытаниям кодеков (или систем), используемых для передачи материалов, созданных в соответствии с Рекомендацией МСЭ-R ВТ.601, в системах доставки и распространения видеоизображения, а также в радиопередающих системах.

Для применений распространения программ характеристики качества могут выражаться в форме субъективного суждения наблюдателей. Так что теоретически такие кодеки могут оцениваться субъективно по настоящим спецификациям. Однако качество кодека, предназначенного для доставки программ, теоретически нельзя определить в форме субъективных параметров, поскольку его выходной сигнал предназначен не для немедленного просмотра, а для последующей студийной обработки, сохранения и/или кодирования с целью дальнейшей передачи. Из-за сложности определения этих параметров качества для разных операций постобработки, в качестве предпочтительного выбран подход указания параметров качества цепочки оборудования, включая функцию постобработки, которая считается репрезентативной для практических применений доставки. Эта цепочка обычно состоит из кодека, за которым следует функция студийной постобработки (или другой кодек в случае оценки качества простой доставки), за которой следует еще один кодек, прежде чем сигнал достигнет наблюдателя. Принятие этой стратегии для спецификации кодеков в применениях доставки означает, что для их оценки также могут использоваться процедуры, приведенные в настоящей Рекомендации.

В области субъективной оценки, где имеется большой опыт, могут быть рекомендованы условия и методики испытаний. Однако следует помнить, что при определении целевых показателей качества или его ухудшения существующие методы позволяют давать не абсолютные субъективные оценки, а, скорее, результаты, на которые в какой-то мере влияет выбор эталонных и/или опорных условий. Для кодеков с фиксированной и с переменной длиной слова, а также для внутрисюжетных и межкадровых кодеков можно использовать одни и те же методики, хотя набор последовательностей испытательных изображений может быть разным.

Наиболее надежный метод оценки для ранжирования высококачественных кодеков заключается в одновременной оценке всех испытываемых систем в одинаковых условиях. Испытания, проведенные независимо, когда речь идет о тонких различиях в качестве, следует использовать для ориентировки, но не в качестве неоспоримого доказательства превосходства.

Полезной субъективной мерой может быть ухудшение качества, определяемое как функция коэффициента ошибок по битам, возникающих в линии передачи между кодером и декодером. В настоящее время недостаточно экспериментальных знаний об истинной статистике ошибок передачи, чтобы рекомендовать параметры модели с учетом кластеров или всплесков ошибок. До появления такой информации для ошибок можно использовать распределение Пуассона.

A1-2 Условия просмотра

Общие условия просмотра для субъективной оценки те же, что приведены в разделе 2 части 1. В следующих разделах приведены конкретные условия просмотра для субъективной оценки цифровых систем.

A1-2.1 Лабораторная среда

Лабораторная среда предназначена для создания условий, необходимых для проверки систем. Конкретные условия просмотра для субъективной оценки в лабораторной среде приведены в таблице 3-1.

ТАБЛИЦА 3-1

Особые условия просмотра для субъективной оценки цифровых систем в лабораторной среде

Условие	Параметр	Значения
a	Отношение расстояния просмотра к высоте изображения	4 <i>H</i> и 6 <i>H</i> ⁽¹⁾
b	Пиковая яркость	70 кд/м ²
c	Угол зрения, охватывающий часть фона, соответствующую спецификациям	≥43° В × 57° Ш
d	Дисплей	Высококачественный экран. Размер ≥ 20 дюймов (50 см) ⁽²⁾

⁽¹⁾ 6 *H* – расчетное расстояние просмотра (DVD) для оценки цифровых систем стандартной четкости, но допустимо и размещение оценщиков на расстоянии 4 *H* при условии, что результаты указываются отдельно.

⁽²⁾ Поскольку имеются некоторые свидетельства того, что размер дисплея может влиять на результаты субъективной оценки, экспериментаторам следует явно указать размер экрана, а также изготавливать и моделировать дисплеи, используемые в любых экспериментах.

A1-2.2 Домашняя среда

Эта среда предназначена для оценки качества цепочки услуг цифрового телевидения со стороны потребителя. Конкретные условия просмотра для субъективной оценки цифрового телевидения стандартной четкости (ТСЧ) в домашней среде приведены в таблице 3-2.

ТАБЛИЦА 3-2

Условия просмотра для субъективной оценки цифровых систем в домашней среде

Условие	Параметр	Значения
a	Отношение расстояния просмотра к высоте изображения	6 Н
b	Размер экрана для формата 4/3	От 25 до 29 дюймов ⁽¹⁾
c	Размер экрана для формата 16/9	От 32 до 36 дюймов ⁽¹⁾
d	Стандарт дисплея	ТСЧ
e	Пиковая яркость	200 кд/м ²
f	Внешняя освещенность на экране (свет от внешних источников, падающий на экран, измеряется перпендикулярно к экрану)	200 люкс

⁽¹⁾ Этот размер экрана удовлетворяет правилам предпочтительного расстояния просмотра (PVD) для PVD = 6 Н.

A1-3 Методы оценки

A1-3.1 Оценка базового качества изображения

Когда кодек оценивается для применений распространения программ, определяемое качество относится к изображениям, декодированным после одного прохода через пару кодер-декодер. Для кодеков доставки программ базовое качество может оцениваться после нескольких последовательных кодеков, чтобы имитировать типичное применение доставки.

В тех случаях, когда оцениваемый диапазон качества невелик, как это обычно бывает для телевизионных кодеков, в качестве методики испытания используется вариант II непрерывной шкалы качества с двумя источниками воздействия, описанный в настоящей Рекомендации. В качестве эталонного условия используется первоначальная исходная последовательность. Дальнейшее рассмотрение посвящено продолжительности представляемых последовательностей. В недавних испытаниях кодеков для компонентного видеоизображения формата 4:2:2 было сочтено целесообразным изменить представление по сравнению с тем, что указано в настоящей Рекомендации. В качестве дополнительного эталона использовались композитные изображения, чтобы обеспечить более низкий уровень качества, по которому можно судить о качестве кодека.

При оценке рекомендуется использовать по крайней мере шесть последовательностей изображений плюс дополнительная последовательность для предварительного обучения. В контексте рассматриваемого применения со снижением скорости передачи данных последовательности должны варьироваться от умеренно критичных до критичных.

По всему тексту настоящего Приложения подчеркивается важность испытания цифровых кодеков с критичными последовательностями изображений в контексте снижения скорости передачи данных телевизионного изображения. Поэтому разумно задать вопрос, насколько критична конкретная последовательность изображений для конкретной задачи уменьшения скорости передачи данных или насколько одна последовательность критичнее другой? Простой, но не особо полезный ответ заключается в том, что для разных кодеков «критичность» означает разные вещи. Например, для внутриполевого кодека неподвижное изображение с большим количеством деталей вполне может оказаться критичным, в то время как для межкадрового кодека, способного использовать сходство между кадрами, та же сцена вообще не представляет никаких трудностей. Некоторые последовательности с движущейся текстурой и сложным движением будут критичными для всех классов кодеков, так что последовательности этого типа наиболее полезны в плане создания или выявления. Сложное движение может принимать форму движений, предсказуемых для наблюдателя, но не для алгоритмов кодирования; примером может служить извилистое периодическое движение.

В одном исследовании возможных статистических показателей критичности изображения, таких как корреляционные методы, спектральные методы, методы условной энтропии и т. д., была найдена простая, но полезная мера, основанная на измерении внутриполевой/межкадровой адаптивной энтропии. Этот метод использовался для "калибровки" последовательностей изображений, предлагаемых для использования в испытаниях МСЭ-R кодеков для скоростей передачи данных 34, 45 и 140 Мбит/с, и оказался полезным для выбора используемых последовательностей. Такие измерения

легче всего выполнять, перенеся эти последовательности изображений на компьютеры обработки изображений и подвергнув их анализу с помощью программного обеспечения.

Ниже представлены некоторые общие рекомендации по выбору критичного материала на тот случай, когда эти методы недоступны.

a) *Внутриполевые кодеки с фиксированной длиной слова*

Несмотря на то, что эти кодеки можно и полезно оценивать на неподвижных изображениях, рекомендуется использовать движущиеся последовательности, поскольку на них легче наблюдать процессы шума кодирования и они выглядят реалистичнее в телевизионных применениях. Если при компьютерном моделировании кодеков используются неподвижные изображения, то обработка должна выполняться по всей оцениваемой последовательности, чтобы, например, сохранить временные аспекты любого шума источника. Выбранные сцены должны содержать как можно больше деталей следующих типов: неподвижные и движущиеся текстурированные области (некоторые с окрашенной текстурой); неподвижные и движущиеся объекты с резкими высококонтрастными краями различной ориентации (некоторые – с окрашенными); неподвижные ровные участки серединного серого цвета. По крайней мере одна последовательность в ансамбле должна демонстрировать едва заметный шум источника и по крайней мере одна последовательность должна быть синтетической (то есть сгенерированной компьютером), чтобы в ней не было дефектов камеры, таких как сканирующая апертура или задержка.

b) *Межкадровые кодеки с фиксированной длиной слова*

Все выбранные испытательные сцены должны содержать движение и как можно больше деталей следующих типов: движущиеся текстурированные области (некоторые – цветные); объекты с резкими, высококонтрастными краями, движущиеся в направлении, перпендикулярном этим краям, и с различной ориентацией (некоторые – окрашенные). По крайней мере одна последовательность в ансамбле должна демонстрировать едва заметный шум источника и по крайней мере одна последовательность должна быть синтетической.

c) *Внутриполевые кодеки с переменной длиной слова*

Эти кодеки рекомендуется испытывать с использованием материала движущихся изображений по тем же причинам, что и кодеки с фиксированной длиной слова. Следует отметить, что благодаря кодированию с переменной длиной слова и соответствующей буферной памяти эти кодеки могут динамически распределять пропускную способность процесса кодирования по изображению. Так, например, если половина изображения состоит из чистого неба, для кодирования которого не требуется много битов, то пропускная способность экономится для других частей изображения, которые могут воспроизводиться с высоким качеством, даже если это критичные области. Из этого следует важный вывод: если последовательность изображений должна быть критичной для такого кодека, то каждая часть экрана должна содержать много деталей. Экран должен быть заполнен движущейся и неподвижной текстурой, как можно большим количеством вариаций цвета и объектами с резкими, контрастными краями. По крайней мере одна последовательность в текстурном ансамбле должна демонстрировать едва заметный шум источника и по крайней мере одна должна быть синтетической.

d) *Межкадровые кодеки с переменной длиной слова*

Это наиболее совершенный класс кодеков, и для их нагрузки требуется наиболее сложный материал. Мало того, что каждая часть сцены должна быть заполнена деталями, как в случае внутриполевого кодека с переменной длиной слова, эти детали также должны демонстрировать движение. Более того, поскольку во многих кодеках используются методы компенсации движения, движение должно быть сложным по всей последовательности. Примерами сложных движений могут служить сцены с одновременным масштабированием и панорамированием камеры; сцена на фоне текстурированной или детализированной шторы, трепещущей на ветру; сцена с предметами, вращающимися в трехмерном пространстве; сцены с детализированными объектами, движущимися по экрану с ускорением. Все сцены должны содержать большое количество объектов, движущихся с разными скоростями, с разными текстурами и контрастными краями, а также наполнены разнообразным цветовым содержанием. По крайней мере одна последовательность в испытательном ансамбле должна

демонстрировать едва заметный шум источника, по крайней мере одна – сложное движение камеры, сгенерированное компьютером из естественного неподвижного изображения (чтобы в нем не было шума и задержки камеры), и по крайней мере одна последовательность должна быть полностью сгенерирована компьютером.

A1-3.2 Оценка качества изображения после постобработки

Эта оценка предназначена для того, чтобы можно было принять решение о пригодности кодека для применений доставки в отношении конкретных процессов постобработки, таких как цветное матирование, замедление движения или электронное увеличение. Минимальная конфигурация оборудования для такой оценки – это один проход через испытуемый кодек, за которым следует проверяемый процесс постобработки, а затем зритель. Тем не менее, более репрезентативным для применений доставки может оказаться использование дополнительных кодеков после постобработки.

Используемая методика испытаний – это вариант II метода непрерывной шкалы качества с двумя источниками воздействия. Однако эталонным условием здесь будет исходный сигнал, подвергнутый той же постобработке, что и декодированные изображения. Если считается благоприятным включение эталона более низкого качества, то он также подвергается той же постобработке.

Испытательные последовательности, необходимые для оценки постобработки, подчиняются точно таким же критериям критичности, что и последовательности для других цифровых применений. Однако этого может быть трудно достичь в последовательностях цветовой рирпроекции переднего плана, поскольку они обычно содержат значительную долю синего фона, лишённого деталей.

Из-за практических ограничений, связанных с возможной необходимостью оценки кодека с несколькими постпроцессами, последовательностей испытательных изображений должно быть как минимум три плюс еще одна для демонстрационных целей. Характер последовательностей зависит от изучаемой задачи постобработки, но должен варьироваться от умеренно критичных до критичных в зависимости от замедления передачи телевизионного сигнала и от рассматриваемого процесса. Для оценки медленного движения может подойти частота воспроизведения в 1/10 частоты кадров исходного сигнала.

A1-3.3 Оценка характеристик искажения

При субъективной оценке искажения изображения в кодеках из-за искажения в канале передачи или излучения следует выбрать как минимум пять, но предпочтительно больше коэффициентов ошибок по битам или избранных условий передачи/излучения, приблизительно логарифмически разнесенных и адекватно проверяющих диапазон, в котором искажения кодека нарастают от «незаметных» до «сильно раздражающих».

Возможно, что потребуются оценка кодека при таких коэффициентах ошибок передачи по битам, которые приводят к настолько редким видимым переходным процессам, что нельзя ожидать, что они произойдут в течение 10-секундной испытательной последовательности. Предлагаемое здесь время демонстрации явно не подходит для таких испытаний.

Если для последующего редактирования демонстраций для субъективной оценки нужно сделать запись выходного изображения кодека в условиях достаточно низкого коэффициента ошибок по битам (приводящего к малому количеству видимых переходных процессов в течение 10-секундного интервала), то следует позаботиться о том, чтобы гарантировать использование записи типичного выходного сигнала кодека за длительный период времени.

Ввиду необходимости исследовать качество кодека в некотором диапазоне коэффициентов ошибок передачи по битам практические ограничения приводят к тому, что адекватным, вероятно, будет сочтено использование трех испытательных последовательностей изображений с дополнительной демонстрационной последовательностью. Продолжительность последовательностей должна составлять порядка 10 секунд, но следует отметить, что зрители могут предпочесть продолжительность в 15–30 секунд. В контексте снижения скорости передачи телевизионного сигнала эти последовательности должны варьироваться от умеренно критичных до критичных.

Поскольку испытания охватывают весь диапазон искажений, подходит метод шкалы искажений с двумя источниками воздействия, который и следует применять.

A1-3.4 Характеристики искажения содержания изображения

Общая концепция характеристик искажения изображения приведена в Приложении 1 к части 1. Для применения этой концепции к системам цифрового телевидения стандартной четкости рекомендуется использовать следующую процедуру.

A1-3.4.1 Определение критичности

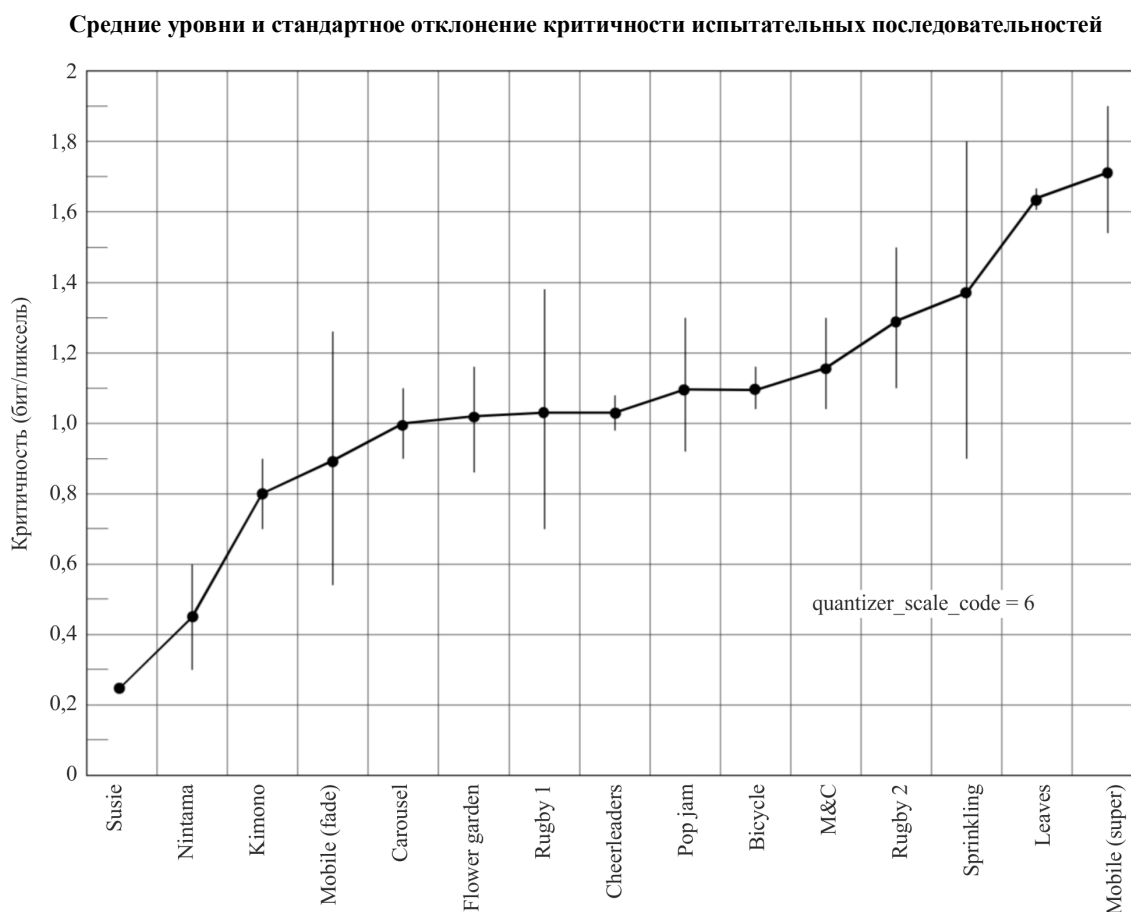
Следует определить специальную меру, называемую "критичностью", которая отражает характеристики испытываемой цифровой телевизионной системы и измеряется путем объективного измерения. В качестве примера системы цифрового телевидения используется система МР@МL MPEG-2 и применяется метод фиксированного квантования критичности на основе энтропии, описанный в Рекомендации МСЭ-R ВТ.1210.

A1-3.4.2 Процедура получения характеристик искажения содержания изображения

- *Шаг 1.* Измерение критичности испытательных последовательностей, используемых при субъективной оценке

Измеряется критичность испытательных последовательностей, используемых для процесса субъективной оценки, описанного в пункте "Шаг 3" ниже. На рисунке 3-1 показано среднее и стандартное отклонение каждой последовательности для примера системы. Большинство последовательностей имеют показатели критичности от 0,8 до 1,4 бит/пиксель. У некоторых последовательностей большое стандартное отклонение, потому что содержание изображения существенно изменяется в течение последовательности.

РИСУНОК 3-1

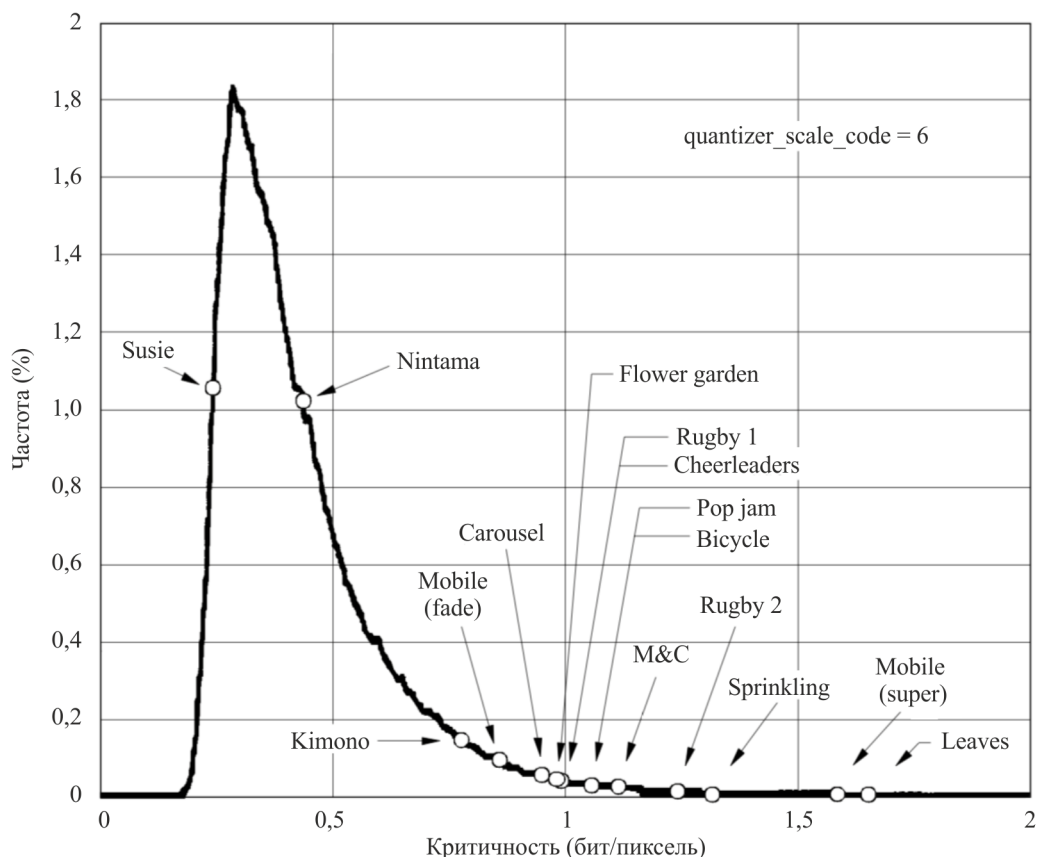


ВТ.0500-03-1

- *Шаг 2.* Измерение распределения критичности телевещательных программ за длительный период времени

Распределение критичности телевещательных программ измеряется в течение длительного периода времени, например, в течение недели. На рисунке 3-2 показан пример распределения, измеренного за одну неделю, в общей сложности 130 часов телевещательных сигналов NTSC, преобразованных для измерения в компонентные сигналы Y/C . Частота проявлений критичности в телевизионных программах рассчитывалась через каждые 5×10^{-3} бит/пиксель. На этом рисунке также показана критичность испытательных последовательностей, используемых для субъективной оценки.

РИСУНОК 3-2
Распределение критичности в телевещательных программах
и критичность испытательных последовательностей



ВТ.0500-03-2

- *Шаг 3.* Проведение субъективной оценки качества изображения испытываемой системы и определение взаимосвязи между критичностью и субъективной оценкой качества изображения

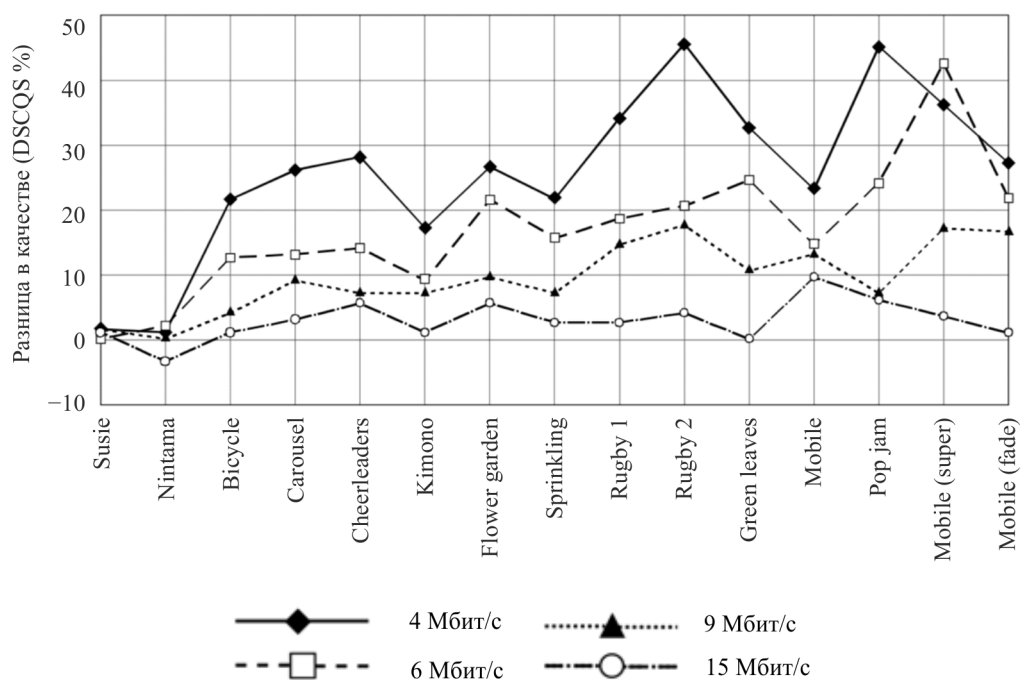
Качество изображения системы цифрового телевидения оценивалось методом DSCQS. Комбинируя результат субъективной оценки с критичностью, полученной на шаге 1, получаем взаимосвязь между критичностью и оценками. На рисунке 3-3 показано качество изображения для примера системы при скоростях передачи 4, 6, 9 и 15 Мбит/с. Значение разницы в качестве на рисунке (DSCQS %) отражает ухудшение по сравнению с эталонной, оригинальной последовательностью компонентов 4:2:2. На рисунке 3-4 показана взаимосвязь между критичностью и разницей в качестве. В этом примере предполагалась линейная зависимость между критичностью и качеством изображения, а линии регрессии получены с использованием метода наименьших квадратов. На рисунке показаны линии регрессии для каждой скорости передачи данных. В общем случае в зависимости от результатов оценки могут применяться нелинейные соотношения.

- *Шаг 4.* Получение характеристики искажения содержания изображения (зависимость качества от частоты появления) путем объединения результатов шага 3 (зависимость качества от критичности) и шага 2 (частота появления критичности).

Комбинируя результаты, полученные на шагах 2 и 3, можно получить характеристики искажения содержания изображения, то есть распределение качества изображения телевизионных программ с цифровым кодированием. Ухудшение качества изображения в программах телевизионного вещания преобразуется в кумулятивную частоту появления. На рисунке 3-5 показаны характеристики искажения содержания изображения для примера системы.

РИСУНОК 3-3

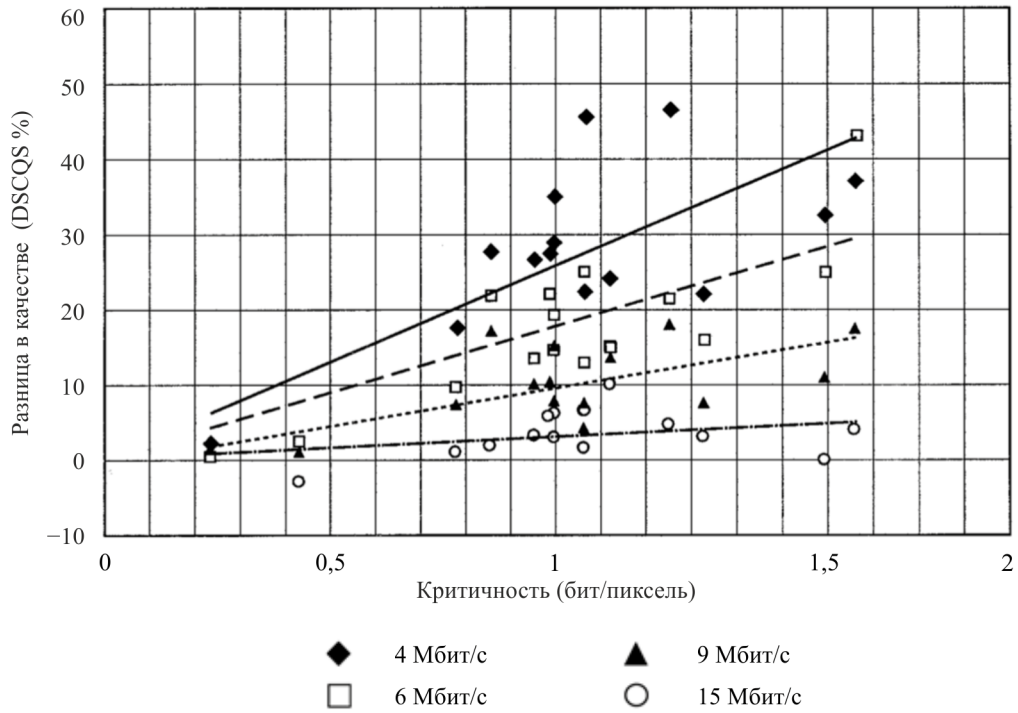
Результаты субъективной оценки (MP@ML при 6Н)



ВТ.0500-03-3

РИСУНОК 3-4

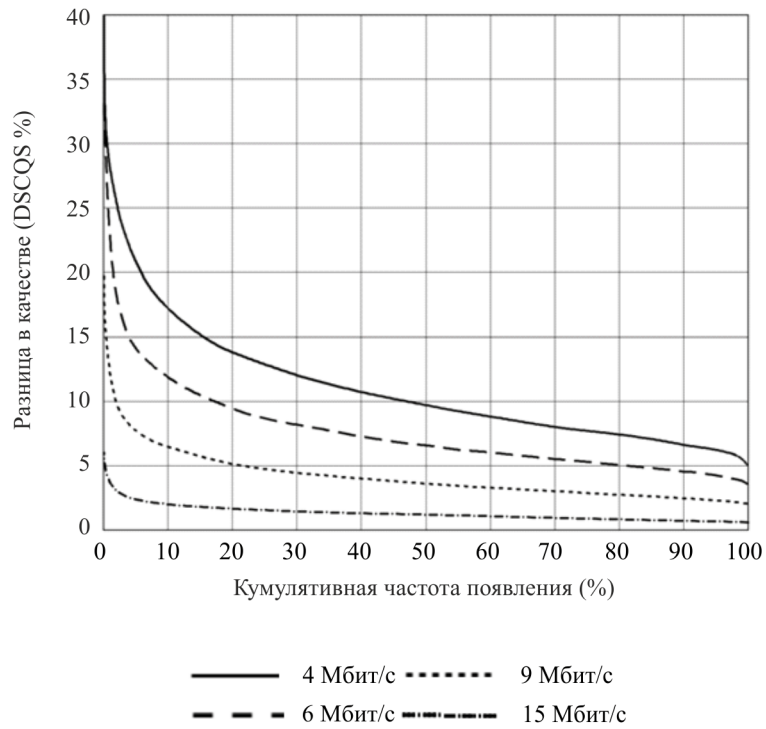
Соотношение между критичностью и оценками (MP@ML при 6Н)



ВТ.0500-03-4

РИСУНОК 3-5

Кумулятивная частота появления искажений изображения (MP@ML на 6Н)



ВТ.0500-03-5

A1-4 Замечания по применению

Если не требуется суждение об абсолютном качестве кодека или вносимом им ухудшении качества, а только порядок ранжирования, или если нужно подтвердить порядок ранжирования, определенный по результатам испытаний с двумя источниками воздействия, то следует использовать метод сравнения парных источников воздействия.

Этот метод, описанный в настоящей Рекомендации, обеспечивает точное сравнение и средства определения соотношения между парами систем. Возможно расширение этого метода для ранжирования нескольких систем по качеству или ухудшению качества. При этом подходе общий порядок ранжирования получается из ранжирования наблюдателями всех возможных пар последовательностей изображений.

Анализ осложняется тем фактом, что наблюдатель может оценивать, например, изображение А лучше, чем изображение В, и изображение В лучше, чем изображение С, но также изображение С лучше, чем изображение А. Это называется "интранзитивной триадой".

Проблема этого метода заключается в том, что количество требуемых демонстраций увеличивается с квадратом числа последовательностей испытательных изображений и кодеков и может стать практически нецелесообразным.

Если широкоэвещательный канал используется для доставки нескольких программ или для масштабируемых или иерархических схем кодирования, то может потребоваться адаптация методик оценки с учетом следующих соображений.

- Критерием приемлемости службы не должна быть прозрачность кодирования исходного изображения; им должна быть способность системы обеспечить реальную альтернативу традиционной службе при выделенной скорости передачи данных. Соответственно, в качестве эталона при испытаниях на качество может оказаться целесообразным использование не материала в несжатой цифровой форме, а материала, доставляемого традиционной системой в типичных условиях приема. Кроме того, может оказаться целесообразным использование материала для испытаний, выбранного для представления диапазона текущего и будущего содержания программ (см. Приложение 3 к части 1). Условия просмотра при испытаниях должны соответствовать указанным в части 1 и в разделе А1-2 настоящего Приложения, а в качестве общего метода испытаний следует использовать метод непрерывной шкалы качества с двумя источниками воздействия (Приложение 2 к части 2).
- Предметом исследования является способность системы поддерживать целостность отдельных потоков программы в условиях полной загрузки канала и ухудшения качества передачи. Соответственно, в испытаниях на ухудшение может оказаться целесообразным обеспечить полную загрузку канала и использовать некоторый диапазон уровней ухудшения, выбранный для отражения диапазона вероятных условий приема (см. Приложение 4 к части 1). Условия просмотра при испытаниях должны соответствовать указанным в части 1 и в разделе А1-2 настоящего Приложения, а в качестве общего метода испытаний следует использовать метод непрерывной шкалы качества с двумя источниками воздействия (см. Приложение 1 к части 2).

ПРИМЕЧАНИЕ. – Когда в одном и том же контексте оцениваются аналоговые и цифровые системы, важно выбрать набор материалов для испытания, сбалансированно отражающий сложность для аналоговых и для цифровых систем. В этом случае может быть полезным применить процедуру многомерного масштабирования для дополнительного анализа.

Приложение 2 к части 3

Субъективная оценка качества изображения телевизионных систем высокой четкости (ТВЧ)

A2-1 Условия просмотра

Если иное не указано в нижеследующей таблице 3-3, то условия просмотра должны быть аналогичны тем, что описаны в разделе 2 части 1.

ТАБЛИЦА 3-3

Условия просмотра для субъективной оценки качества изображения ТВЧ

Условие	Параметр	Значения
a	Отношение расстояния просмотра к высоте изображения	3
b	Пиковая яркость экрана (кд/м ²) ⁽¹⁾	150–250
c	Отношение яркости неактивного экрана к пиковой яркости ⁽²⁾	≤ 0,02
d	Отношение яркости экрана при отображении только уровня черного в абсолютно темном помещении к яркости, соответствующей пиковому уровню белого ⁽³⁾	примерно 0,01
e	Отношение яркости фона за дисплеем к пиковой яркости изображения	примерно 0,15
f	Освещенность, вызванная другими источниками ⁽⁴⁾	слабая
g	Цветность фона	D ₆₅
h	Угол, образованный частью фона, соответствующей приведенной выше спецификации ⁽⁵⁾ . Он должен сохраняться для всех наблюдателей	53° по высоте × 83° по ширине
i	Расположение наблюдателей	В пределах ±30° по горизонтали от центра экрана. Предел по вертикали изучается
j	Размер экрана ⁽⁶⁾	1,4 м (55 дюймов)

⁽¹⁾Пиковая яркость на экране соответствует видеосигналу со 100%-ной амплитудой.

⁽²⁾ На этот параметр может повлиять освещение в помещении, а также диапазон контрастности дисплея.

⁽³⁾ Уровень черного соответствует видеосигналу с амплитудой 0%.

⁽⁴⁾ Освещение в помещении должно быть отрегулировано так, чтобы можно было выполнить условия с и e.

⁽⁵⁾ Рекомендуется минимум 28° по высоте x 48° по ширине.

⁽⁶⁾ Если дисплеи указанного размера недоступны, следует использовать значения ≥76,2 см (30 дюймов). См. Примечание 3 к части 1.

A2-2 Методы оценки

Субъективные оценки общего качества изображения ТВЧ, передаваемого радиовещательной системой, следует проводить с использованием метода непрерывной шкалы качества с двумя источниками воздействия (Приложение 2 к части 2) с использованием в качестве эталона изображения ТВЧ студийного качества.

Оценку характеристик искажения радиовещательной системы ТВЧ следует проводить с использованием шкалы искажений с двумя источниками воздействия (Приложение 1 к части 2) с использованием в качестве эталона студийного изображения ТВЧ или изображения, переданного без искажений.

Когда предметом исследования является качество изображения при содержании программ и условиях передачи, близких к встречающимся на практике, следует рассмотреть описание характеристики общих искажений, приведенное в Приложении 4 к части 1.

При использовании этих методов необходимо следить за тем, чтобы отличить влияние формата дисплея от влияния формата базовой системы (например, любого преобразования с повышением частоты). Для учета разных форматов дисплея можно выполнить дополнительные оценки с использованием разных дисплеев, если это уместно и целесообразно.

Некоторые системы трансляции ТВЧ могут включать в себя встроенный обычный телевизионный формат (для обратной совместимости). Таким образом, существует необходимость оценки адекватности обычных телевизионных изображений, встроенных в радиовещательный сигнал ТВЧ, с точки зрения качества изображения. Для этих систем следует применять условия просмотра и методы оценки, приведенные в Приложении 1 к части 3.

К цифровым системам трансляции ТВЧ, в которых используются схемы понижения скорости передачи, следует применять базовые концепции и процедуры, описанные в Приложении 1 к части 3.

A2-3 Испытательные материалы

В Отчете МСЭ-R ВТ.2245 представлен широкий спектр неподвижных и движущихся изображений. Их предпочтительно использовать в качестве общих материалов для испытания для оценки качества ТВЧ.

Приложение 3 к части 3

Субъективная оценка качества буквенно-цифровых и графических изображений в телетексте и аналогичных текстовых службах

Введение

Существуют системы, которые обрабатывают графические и буквенно-цифровые изображения и передают их с помощью соответствующих цифровых кодов. Буквенно-цифровые и графические изображения имеют специфический характер, отличный от обычного телевизионного изображения, и умственный процесс, связанный с их субъективной оценкой, может отличаться.

В настоящей Рекомендации предлагаются методы субъективной оценки качества изображений, содержащихся в современных телевизионных программах. Необходимы исследования в области качества буквенно-цифровых и графических изображений, передаваемых по телевизионным каналам и используемых для нескольких новых служб, в которых для описания буквенно-цифровых и графических изображений применяются цифровые коды. На качество изображения влияют несколько параметров передачи: разрешение страницы (количество строк на страницу и символов в строке) при альфа-мозаичном кодировании телетекста, разрешение знакоместа (количество пикселей и строк на знакоместо) при кодировании с динамически переопределяемым набором символов (DRCS, см. Рекомендацию МСЭ-R ВТ.653) и разрешение изображения при использовании широковещательной аудиографии, факсимиле или телетекста. Кроме того, следует учитывать влияние ошибок передачи, которые могут повлиять на коды. Таким образом, необходимы измерения качества и определения отношений объективных и субъективных оценок для этих параметров.

Исследования показали, что для оценки качества этих изображений, характеристики которых могут отличаться от характеристик обычных телевизионных изображений, требуется учитывать иные аспекты. На различные атрибуты качества – разборчивость, качество, комфорт, раздражение, усилие при чтении, утомление и эстетические соображения – влияют такие параметры, как формат пикселя, разрешение знакоместа, интервалы, цвета и верстка. Здесь рассматриваются три основных аспекта: условия просмотра, методы оценки и контекст оценки.

Ввиду важности создания основы для субъективных оценок качества буквенно-цифровых и графических изображений во всех отчетах о результатах испытаний должно быть представлено по возможности самое полное описание конфигурации испытаний, материалов для испытания, наблюдателей и методов.

А3-1 Условия просмотра

В части 1 определены условия просмотра телевизионных изображений, соответствующие низким уровням освещенности в помещении. Вероятно, что буквенно-цифровые и графические изображения будут просматриваться также и при нормальных условиях освещения. Таким образом, для изучения предлагается дополнительный набор условий просмотра: освещенность 500 люкс, максимальная яркость экрана от 70 до 200 кд/м², коэффициент контрастности экрана от 30 до 50 и значение отношения яркости фона (стен помещения) к максимальной яркости экрана 1/4. Также следует учитывать расстояния просмотра от четырех- до восьмикратной высоты изображения.

А3-2 Методы оценки

По шрифтовым аспектам проведено большое количество исследований. В большинстве из них используются "показатели эффективности", такие как пороги узнавания или распознавания, коэффициент распознавания, скорость чтения и т. д. "Субъективные показатели", обычно используемые при оценке качества телевизионных изображений, применяются в очень немногих исследованиях. Считается, что новые системы, передаваемые по телевизионным каналам, должны иметь хорошие характеристики (например, процент успешного распознавания букв выше 95%). Таким образом, могут эффективно использоваться шкалы качества и искажения, приведенные в настоящей Рекомендации, хотя необходимы исследования, чтобы установить, каким образом эти шкалы можно связать с удобочитаемостью. Была проведена попытка сравнения с методами оценки качества речи (МСЭ-Т) и предложена пятибалльная шкала "усилия при чтении" для дальнейшего изучения.

В другом методе сравниваются результаты субъективных оценок, выполненных с использованием двух разных пятибалльных шкал, приведенных в таблице 3-4.

ТАБЛИЦА 3-4

Шкалы разборчивости и затрачиваемых усилий при чтении

Шкала разборчивости	Шкала затрачиваемых усилий при чтении
Отличная разборчивость	Никаких усилий при чтении
Хорошая разборчивость	Необходимо внимание, но заметные усилия при чтении отсутствуют
Удовлетворительная разборчивость	Умеренные усилия при чтении
Плохая разборчивость	Значительные усилия при чтении
Неприемлемая разборчивость	Чрезвычайные усилия при чтении

Было признано, что важно очень четко сформулировать каждую степень оценки. Средние значения оценок, полученных с помощью шкалы затрачиваемых при чтении усилий обычно выше средних значений, полученных с помощью шкалы разборчивости, а диапазон оценок, данных наблюдателями, шире в случае шкалы затрачиваемых при чтении усилий.

В другом эксперименте для оценки мнений как об общем качестве, так и об общей разборчивости машинописного текста, передаваемого телевизионной системой с переменными стандартом строк и шириной полосы пропускания, использовалась шкала качества, описанная в пункте А3-4.1 части 2. Для каждого мнения были найдены две модели, одна из которых более сложная и точная, но в обеих используется идея добавления "шкалы ухудшения", отражающей совокупное влияние ограниченного определения по горизонтали и вертикали. Разборчивость также измерялась как величина, пропорциональная числу правильно распознанных символов. Однако в таких случаях разборчивость остается высокой даже при низком качестве изображения, и очевидно, что первый критерий обычно менее полезен.

В другом исследовании проводилось сравнение объективных и субъективных методов на печатном текстовом материале с использованием символов фиксированной и переменной ширины. Субъективные методы оказались более чувствительными. Такое же исследование было повторено с использованием дисплея на основе электронно-лучевой трубки, на этот раз с применением только субъективных методов. Использование этих субъективных методов привело к результатам, связанным с визуально оптимальными размерами фиксированных и переменных матриц.

A3-3 Контекст оценки

В новом подходе к оценке службы рассматривается случай, когда действия пользователя могут быть точно определены. Оценки получают не традиционным методом представления изображений, когда зрители просто дают стандартные субъективные оценки. Вместо этого зрители используют представленные изображения так, как если бы они пользовались исследуемой службой, и все оценки выполняются в этом контексте.

Эмуляция использования службы не исключает применения обычных субъективных мер. Тем не менее, она создает контекст для субъективных оценок, наиболее подходящий для исследуемой службы. Это также позволяет использовать объективные показатели квалификации зрителей и разрабатывать новые субъективные показатели, наиболее подходящие для исследуемых служб и параметров. Наконец, это создает более надежную основу для обобщения оценок, полученных в лаборатории, с оценками в условиях эксплуатации.

Приложение 4 к части 3

Субъективная оценка качества изображения в многопрограммных службах⁵

Введение

Для субъективной оценки качества отдельных программ, сжатых и закодированных с постоянной битовой скоростью (CBR) в многопрограммной службе, следует использовать процедуры субъективной оценки, описанные в Приложениях 1 или 2 к части 3, и процедуру, описанную в пункте A4-2 настоящего Приложения.

Для субъективной оценки качества отдельных программ, сжатых и закодированных с переменной битовой скоростью (VBR) с применением таких методов, как статистическое мультиплексирование или совместное кодирование, в многопрограммной службе следует использовать процедуры субъективной оценки, описанные в Приложениях 1 или 2 к части 3, и процедуру, описанную в пункте A4-3 настоящего Приложения.

A4-1 Общие рекомендации по оценке

- Оценки качества тематических каналов следует проводить с использованием материала для испытаний, схожего по содержанию и критичности с тем, который обычно передается по этим каналам.
- Чтобы оценить общее воспринимаемое качество программы, "мгновенное" качество которой изменяется в течение определенного периода времени, следует использовать процедуры, описанные в пунктах A4-2 и A4-3.
- При испытаниях, в которых многопрограммные службы сравниваются с низкокачественным материалом, следует применять и дополнительно изучать масштабирование результатов для систем с низкими эталонами качества в соответствии с замечаниями, содержащимися в описании метода DSCQS.

⁵ Включает термин "Статистическое мультиплексирование", или службы Stat-Mux.

A4-2 Процедуры субъективной оценки изображений для многопрограммных служб с постоянной скоростью передачи битов

Субъективную оценку качества изображения каждой программы ТСЧ и ТВЧ можно проводить независимо с использованием методов, описанных в Приложении 1 (ТСЧ) или Приложении 2 (ТВЧ) к части 3. Для оценки базового качества системы следует использовать общий метод испытаний DSCQS (описанный в Приложении 2 к части 2). Для оценки программ с ухудшением качества передачи следует использовать общий метод испытаний DSIS (описанный в Приложении 1 к части 2).

A4-3 Процедуры субъективной оценки изображений для многопрограммных служб с переменной скоростью передачи битов

Субъективную оценку качества изображения программ ТСЧ и ТВЧ с VBR-кодированием можно выполнять с использованием методики DSCQS. Следует также обратить внимание на выбор материалов для испытания, поскольку качество изображения может зависеть от содержания изображения всех мультиплексируемых программ.

Приложение 5 к части 3

Экспертная оценка качества изображения систем цифрового воспроизведения видеоизображения для большого экрана⁶ в кинотеатрах

A5-1 Введение

В прошлые годы просмотр экспертами часто использовался для быстрой проверки характеристик общего процесса видеопроизводства.

В этом Приложении описан метод испытаний с просмотром экспертами, гарантирующий согласованность результатов, полученных в разных лабораториях, при использовании ограниченного числа экспертов-оценщиков.

A5-2 Почему новый метод основан на "просмотре экспертами"

Полезно указать на преимущества, вытекающие из применения предложенной методики.

Во-первых, для формальных испытаний с целью субъективной оценки обычно требуется как минимум 15 наблюдателей, выбранных в качестве "неспециалистов", а для этого необходимы длительные испытания и постоянный поиск новых наблюдателей. Такое количество наблюдателей требуется для достижения необходимой чувствительности, с тем чтобы испытываемые системы можно было надежно дифференцировать и ранжировать или использовать в качестве надежно оцененного эквивалента.

Во-вторых, при использовании наблюдателей-неспециалистов традиционные испытания могут не выявить различий, которые при длительном просмотре становятся заметными даже специалистам.

В-третьих, традиционные оценки обычно устанавливают показатели качества (или различия в качестве), но не определяют непосредственно артефакты или другие физические проявления, приводящие к этим показателям.

Предлагаемая здесь методика призвана решить все три проблемы.

⁶ Цифровые системы воспроизведения видеоизображения для большого экрана (LSDI) – это семейство цифровых систем обработки изображений, применимых для таких программ, как кинофильмы, сценические представления, спортивные мероприятия, концерты, культурные мероприятия и т. д., от съемки до демонстрации на большом экране с высоким разрешением в надлежащем образом оборудованных кинотеатрах, концертных залах и в других крупных аудиториях.

A5-3 Определение экспертов

Для целей настоящего Приложения "зритель-эксперт" – это человек, знакомый с материалом, используемым для оценки, который "знает, на что смотреть" и, возможно, хорошо разбирается в деталях алгоритма, используемого для обработки оцениваемого видеоматериала. В любом случае, "зритель-эксперт" – это человек с большим опытом работы в области исследования качества, профессионально занимающийся конкретной областью, к которой относится испытание. Например, при организации сеанса просмотра экспертами материала LSDI следует выбирать специалистов по кинопроизводству или постпроизводству или по производству высококачественного видеоконтента (кинооператоры, цветокорректоры и т. д.); этот выбор должен производиться с учетом их способности давать оригинальную субъективную оценку качества изображения LSDI и артефактов, вызываемых сжатием.

A5-4 Выбор оценщиков

Просмотр экспертами – это сеанс оценки, основанный на мнениях оценщиков, которые высказывают суждения о визуальном качестве и/или ухудшении видимости.

Основная группа экспертов состоит из пяти-шести человек. Столь небольшая численность группы облегчает подбор оценщиков и ускоряет процесс принятия решений.

В соответствии с потребностями эксперимента допустимо использовать несколько базовых групп экспертов, составляющих более крупную объединенную группу (например, из разных лабораторий).

Признано, что когда эксперты испытывают свою собственную технологию, они могут склоняться к необъективности оценок, поэтому следует избегать включения лиц, непосредственно вовлеченных в разработку испытываемой системы (систем).

Все оценщики должны пройти отбор на нормальную или скорректированную до нормальной остроту зрения (испытание Снеллена) и нормальное цветовое зрение (испытание Ишихары).

A5-5 Испытательный материал

Испытательные материалы выбираются так, чтобы они отражали диапазон производственных значений и уровней сложности предполагаемого реального контекста, в котором будет использоваться испытываемая система (системы). При выборе предпочтение отдается более сложному материалу без чрезмерной экстремальности. В идеале следует использовать пять-семь испытательных последовательностей.

Способ выбора материала также может различаться в зависимости от предполагаемой сферы применения испытываемой системы.

Поэтому здесь не дается никаких дополнительных указаний относительно правил выбора материала для испытаний; в связи с вышеизложенными соображениями решение остается за разработчиком испытания.

A5-6 Условия просмотра

Условия просмотра должны быть полностью описаны в протоколе испытаний и соответствовать таблице 3-5, оставаясь постоянными в течение испытания.

ТАБЛИЦА 3-5

Обзор условий просмотра

Условия просмотра	Значения параметров	
	Минимальное	Максимальное
Размер экрана (м)	6	16
Расстояние просмотра ⁽¹⁾	1,5 Н	2 Н
Яркость проектора (в центре экрана, пиковая яркость)	34 кд/м ²	48 кд/м ²
Яркость экрана (проектор выключен)		< 1/1000 яркости проектора

⁽¹⁾ Когда расстояние просмотра ближе к 1,5 Н, следует использовать метод демонстрации "бабочка". Если используется метод демонстрации "бок о бок", то расстояние просмотра должно быть ближе к значению 2 Н.

А5-7 Методика

А5-7.1 Сеансы оценки

Каждый сеанс оценки (определяемый как набор испытательных заданий для данной группы наблюдателей) должен состоять из двух этапов (этапа I и этапа II).

А5-7.1.1 Этап I

Этап I состоит из формального субъективного испытания, проводимого в контролируемой среде (см. пункт А5-6), который позволяет получить достоверные, точные и воспроизводимые результаты. Эксперты индивидуально оценивают показанный материал, используя описанную ниже шкалу оценки. Членам группы не разрешается обсуждать то, что они видят, или управлять демонстрацией. На этом этапе эксперты не должны знать испытываемую схему кодирования и порядок представления материала для испытаний. Во избежание необъективности оценки предназначенный для испытания материал демонстрируется в случайном порядке.

А5-7.1.1.1 Демонстрация материала

Метод демонстрации объединяет элементы метода непрерывной оценки качества с одновременным воздействием двух источников (SDSCE) (Приложение 6 к части 2) и метода двух источников воздействия с непрерывной шкалой качества (DSCQS) (Приложение 2 к части 2). Для ссылок его можно назвать методом одновременного двойного воздействия (SDS).

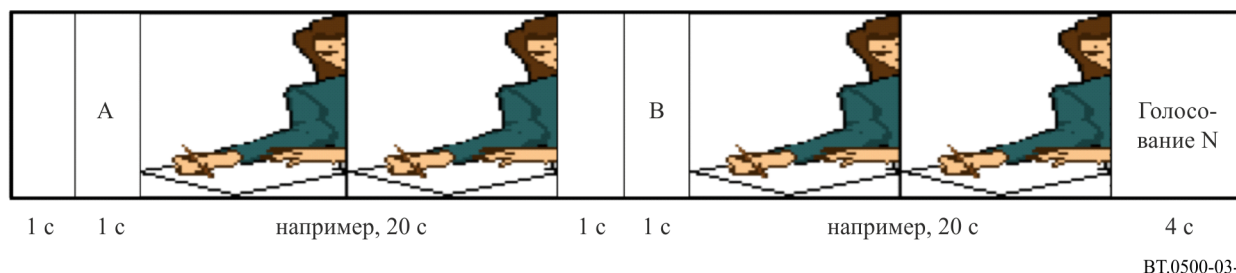
Как и в случае метода SDSCE, каждое испытание включает в себя демонстрацию на разделенном экране материала двух изображений. В большинстве случаев одним из источников изображения служит эталон (то есть исходное изображение), а другим – испытательное изображение; в других случаях, оба изображения получают из эталонного изображения. Эталоном служит исходный материал, представленный прозрачно (то есть не подвергаемый никакому сжатию, кроме того, которое подразумевается для исходного носителя записи). Испытательный материал должен представлять собой исходный материал, обработанный одной из испытываемых систем. Скорость передачи битов и/или уровень качества должны соответствовать требованиям, указанным в плане испытаний. В отличие от испытаний методом SDSCE, наблюдатели не знают об условиях, соответствующих двум компонентам пары изображений.

Демонстрация на разделенном экране выполняется с использованием традиционного разделенного экрана без зеркалирования или методом "бабочка", когда изображение на правой стороне экрана обращено по горизонтали. Поскольку используются изображения полной ширины, одновременно может отображаться только половина каждого изображения. При каждой демонстрации на каждой стороне экрана отображается одна и та же половина изображения.

Как и при использовании метода DSCQS, пара изображений демонстрируется два раза подряд, один раз для ознакомления и рассматривания, а второй – для проверки впечатления и оценки. Каждая последовательность длится 15–30 с. В помощь оценщикам каждая последовательность может быть помечена в начале фрагмента (см. пример разделенного экрана без зеркалирования на рисунке 3-6).

РИСУНОК 3-6

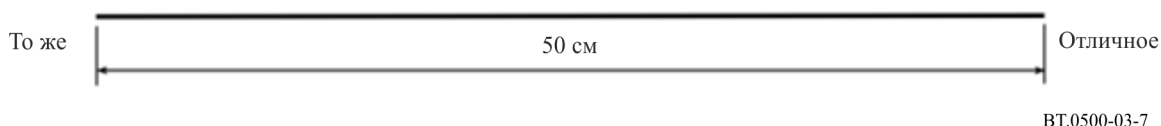
Пример разделения экрана без зеркалирования



A5-7.1.1.2 Шкала оценки

Критерием приемлемости в применениях LSDI служит неотличимость испытуемого (то есть сжатого) изображения от эталонного. Для оценки испытуемых систем можно использовать несколько широко применяемых методов. Предлагаемый метод – рекомендуемые шкалы сравнения воздействий (см. Приложение 4 к части 2). Конкретным примером шкалы может служить некатегорическая (непрерывная) шкала ТО ЖЕ – ОТЛИЧНОЕ, описанная в пункте А4-4.2 Приложения 4 к части 2.

РИСУНОК 3-7



A5-7.1.1.3 Сеанс оценки

Сеанс, который может включать в себя нескольких демонстраций в зависимости от количества условий испытаний, состоит из испытаний двух типов: пробного и контрольного. В процессе пробного испытания на одной половине экрана демонстрируется эталонное изображение, а на другой – проверяемое. В ходе контрольного испытания на обеих половинах экрана демонстрируется эталонное изображение. Целью контрольного испытания является оценка степени необъективности суждения.

Для каждой испытуемой системы требуются следующие пробные испытания на каждой испытательной последовательности.

ТАБЛИЦА 3-6

Левая часть экрана	Правая часть экрана
Левая половина эталонного изображения	Левая половина испытательного изображения
Правая половина эталонного изображения	Правая половина испытательного изображения
Левая половина испытательного изображения	Левая половина эталонного изображения
Правая половина испытательного изображения	Правая половина эталонного изображения

Предпочтительно, чтобы в каждом из указанных выше случаев производилось по меньшей мере два повтора. Для каждой испытуемой системы требуются следующие контрольные испытания на каждой испытательной последовательности.

ТАБЛИЦА 3-7

Левая часть	Правая часть
Левая половина эталонного изображения	Левая половина эталонного изображения
Правая половина эталонного изображения	Правая половина эталонного изображения

Опять же предпочтительно, чтобы в каждом из указанных выше случаев производилось по меньшей мере два повтора.

Пробный сеанс следует разделить на демонстрации продолжительностью не более одного часа, разделенные 15-минутными перерывами для отдыха. Пробные и контрольные испытания, образуемые в результате комбинации кодеков и испытательных последовательностей, распределяются по демонстрациям в псевдослучайном порядке. Полезно наложить на этот процесс некоторые ограничения, хотя это и усложнит его. Например, если производятся четыре демонстрации, то можно случайным образом устанавливать очередность четырех пробных испытаний для данного кодека и испытательной последовательности в ходе одной из демонстраций. Преимущество этого подхода состоит в том, что пробные испытания каждой системы распределяются по всему сеансу испытаний.

A5-7.1.1.4 Обработка результатов испытаний

Для данного пробного испытания пробная оценка – это расстояние от начальной точки "ТО ЖЕ" шкалы до отметки на шкале 0–100, сделанной наблюдателем. Результаты анализируются с определением показателя усредненного мнения (MOS), который используется для ранжирования испытуемых систем. В зависимости от количества наблюдений на систему (число наблюдателей × число испытательных последовательностей × число повторов) данные могут подвергаться дисперсионному анализу (ANOVA)⁷. Результаты контрольных испытаний могут использоваться для определения базового разброса при "случайных" суждениях.

A5-7.1.2 Этап II

Одной из основных целей этапа II является уточнение относительного ранжирования по результатам этапа I, точность и надежность которого могут быть снижены из-за ограниченного числа наблюдателей и/или пробных суждений. Еще одна важная задача – выявление характеристик восприятия различий между изображениями, на которых основываются суждения, вынесенные на первом этапе.

Эта часть состоит из рассмотрения группой экспертов показанного материала. Теперь экспертам разрешается обсуждать представленный материал, повторять его частично или полностью столько раз, сколько необходимо для рассмотрения и/или демонстрации, и прийти к согласованному суждению и описанию того, что они видят. По запросу экспертов разрешаются такие режимы, как замедленное или покадровое воспроизведение, стоп-кадр и т. п. Для этого потребуется некоторое взаимодействие и вмешательство со стороны администратора испытаний.

A5-7.1.2.1 Группирование материала для испытания

Для правильного выполнения этапа II испытаний необходимо сгруппировать выбранный для испытания материал по содержанию, получив так называемый базовый набор для просмотра экспертами (BES), в котором сгруппированы все кодированные последовательности, полученные из одной и той же исходной последовательности, расположенные в соответствии с рейтингом, полученным на этапе I.

Испытательный материал располагается в порядке от наименьшего значения MOS до наибольшего. Количество BES равно числу испытательных последовательностей.

A5-7.1.2.2 Тренировочный сеанс просмотра экспертами

Тренировочный сеанс просмотра экспертами (BEV) – это сеанс обсуждения, в ходе которого эксперты изучают весь материал, включенный в BES; одна из задач состоит в том, чтобы подтвердить или изменить порядок ранжирования, полученный в результате этапа I формальных испытаний. Следовательно, должна быть подтверждена или изменена относительная видимость различий.

⁷ Для применения методов инференциальной статистики, таких как ANOVA, достаточно в общей сложности 10–20 наблюдений в наихудших условиях, представляющих интерес.

A5-7.1.2.3 План этапа II

На этапе II выполняются все BEV. Экспертам известно, что порядок представления является результатом ранжирования на этапе I, но не известно о какой-либо связи между ранжированием систем и теми, кто имеет к ним отношение.

Этап II проводится как групповое мероприятие, приводящее к консенсусу оценщиков.

Перед началом второго этапа с оценщиками проводится инструктаж, возможно, в текстовом формате, решающий следующие задачи:

- ознакомление с материалом каждого BEV;
- обсуждение рейтинга материалов в каждом BEV; если группа не согласна с рейтингом, определяется новый рейтинг;
- комментирование каждого случая с выражением подробных замечаний о характере наблюдаемых различий, если таковые имеются;
- документирование рейтингов, замечаний и наблюдений.

Администратор испытаний собирает все комментарии групп и проверяет расхождения между ними. Во время испытаний результаты этапов I и II отдельных групп хранятся в тайне во избежание их влияния на последующие группы. По возможности администратор испытаний может выявлять расхождения и поддерживать разрешение путем дальнейших испытаний при спорных рейтингах. Целью этого последнего шага является достижение общего консенсуса.

A5-8 Отчет

За окончательный отчет об испытаниях отвечает администратор испытаний.

В этом отчете должна быть представлена следующая информация:

- результаты этапа I (включая таблицы MOS, а также в соответствующих случаях результаты статистического анализа);
- замечания экспертов, собранные на этапе II;
- комментарии к любым переоценкам рейтингов;
- любая соответствующая информация по условиям просмотра, характеристикам входного сигнала, обработке сигналов, характеристикам видеопроектора, настройке видеопроектора, цветности, выбору зрителей и условиям испытаний;
- полная характеристика воспроизводящего устройства (средняя наработка на отказ и т. д.);
- резюме и выводы.

Приложение 6 к части 3

Субъективная оценка качества изображения в мультимедийных приложениях

A6-1 Введение

Во многих странах началось развертывание систем цифрового телевизионного вещания, которые позволяют доставлять потребителям мультимедийные программы и данные, содержащие видео- и аудиозаписи, фотографии, текст и графику.

Для определения требований к качеству и проверки пригодности технических решений, рассматриваемых для каждого применения, необходимы стандартизированные методы субъективной оценки. Субъективные методики необходимы потому, что они позволяют производить измерения, благодаря которым производители могут непосредственно прогнозировать реакцию конечных пользователей.

Система ТВ-вещания, необходимая для доставки мультимедийных программ, заметно отличается от используемой в настоящее время: доступ к информации осуществляется через стационарные и/или мобильные приемники; частота кадров может быть фиксированной или переменной; возможен широкий диапазон размеров изображения (от SQCIF до ТВЧ); видеоизображение обычно связано со звуковым сопровождением, текстом и/или звуком; видеоизображение может обрабатываться с помощью передовых видеокодеков; а предпочтительное расстояние просмотра в большой мере зависит от применения.

В этом новом контексте следует применять методы субъективной оценки, описанные в части 2. Кроме того, может проводиться исследование мультимедийных систем с использованием новых методик для удовлетворения требований, предъявляемых пользователями к характеристикам мультимедийных программ.

Это Приложение посвящено неинтерактивной субъективной оценке качества видеоизображения в мультимедийных применениях. Эти методы могут применяться для разных целей, включая, помимо прочего, выбор алгоритмов, оценку характеристик аудиовизуальных систем и оценку уровня качества видеоизображения во время сеанса аудиовизуального соединения.

А6-2 Общие особенности

А6-2.1 Условия просмотра

Рекомендуемые условия просмотра приведены в таблице 3-8. Следует использовать дисплей подходящего размера и типа для исследуемого применения. Поскольку в мультимедийных применениях используются разные технологии дисплеев, следует сообщить всю необходимую информацию о дисплее, используемом при оценке (производитель, модель, спецификации и т. д.).

Когда для представления последовательностей используются системы на базе ПК, следует также сообщить характеристики этих систем (например, тип видеокарты).

В таблице 3-9 приведен пример записи данных о конфигурации испытуемой мультимедийной системы.

Если испытательные изображения получены с использованием определенной комбинации декодера и устройства воспроизведения, то в целях анонимного воспроизведения изображения должны отделяться от фирменной оболочки. Это необходимо для того, чтобы на оценку качества не влияло знание исходной среды.

Если в оцениваемых системах используется уменьшенный формат изображения, такой как CIF, SIF, QCIF и т. д., то последовательности должны отображаться в окне экрана дисплея. Фон на экране должен быть серого цвета (50%).

ТАБЛИЦА 3-8

Рекомендуемые условия просмотра для оценки качества мультимедиа

Параметр	Значение
Расстояние просмотра ⁽¹⁾	Ограниченное: 1–8 Н Неограниченное: в зависимости от предпочтений зрителя
Пиковая яркость экрана	70–250 кд/м ²
Отношение яркости неактивного экрана к пиковой яркости	≤ 0,05
Отношение яркости экрана при отображении только уровня черного в абсолютно темном помещении к яркости, соответствующей пиковому уровню белого	≤ 0,1
Отношение яркости фона за дисплеем изображения к пиковой яркости изображения ⁽²⁾	≤ 0,2
Цвет фона ⁽³⁾	D65
Освещенность фонового помещения ⁽²⁾	≤ 20 люкс

(1) В общем случае расстояние просмотра зависит от применения.

(2) Это значение, обеспечивающее максимальную вероятность обнаружения искажений; для некоторых применений допускаются более высокие значения, или же они определяются применением.

(3) Для дисплеев ПК цвет фона должен быть максимально приближен к цвету "белой точки" на экране.

ТАБЛИЦА 3-9

Конфигурация испытываемой мультимедийной системы

Параметр	Описание
Тип дисплея	
Размер экрана	
Видеокарта	
Производитель	
Модель	
Сведения об изображении	

А6-2.2 Исходные сигналы

Источник сигнала обеспечивает непосредственно эталонное изображение и входной видеосигнал для испытываемой системы. Исходных последовательностей должно быть как можно больше. Как правило, видеосигнал записывается в виде мультимедийных файлов с использованием формата YUV (4:2:2, 4:4:4) или RGB (24 или 32 бита). Когда экспериментатору нужно сравнить результаты разных лабораторий, необходимо использовать общий набор исходных последовательностей, чтобы исключить дополнительный источник разброса.

А6-2.3 Подбор материалов для испытания

Решающее значение для интерпретации результатов субъективной оценки имеют количество и тип испытательных сцен. Некоторые процессы могут приводить к одинаковому ухудшению качества большинства последовательностей. В таких случаях значимую оценку могут давать результаты, полученные при небольшом количестве последовательностей (например, двух). Однако новые системы часто подвержены влиянию, сильно зависящему от сцены или содержания последовательности. В таких случаях количество и тип испытательных сцен следует выбирать так, чтобы обеспечить разумное обобщение до уровня обычных программ. Кроме того, следует выбрать материал, "критичный, но не чрезмерно" для испытываемой системы. Под фразой "не чрезмерно" подразумевается, что сцену можно считать частью обычного содержания телевизионной программы. Полезным

признаком сложности сцены могут служить характеристики пространственно-временного восприятия. Измерение характеристик пространственно-временного восприятия подробно рассматривается в Приложении 6 к части 1.

А6-2.4 Диапазон условий и привязки

Поскольку большинство методов оценки чувствительны к изменениям диапазона и распределения наблюдаемых условий, в сеансы оценки следует включать весь диапазон переменных факторов. Однако его можно аппроксимировать, опираясь на более ограниченный диапазон, представив также некоторые условия, соответствующие крайним значениям шкал. Их можно представить в качестве примеров и определить как экстремальные (прямая привязка) или распределить по всему сеансу и не определять в качестве экстремальных (косвенная привязка). По возможности следует использовать широкий диапазон показателей качества.

А6-2.5 Наблюдатели

После отбора должно оставаться не менее 15 наблюдателей. Они должны быть неспециалистами, то есть их обычная работа не должна иметь прямого отношения к качеству изображения, и они не должны быть опытными оценщиками. Перед сеансом наблюдателей следует проверить на нормальную (скорректированы до нормальной) остроту зрения по таблице Снеллена или Ландольта и на нормальное цветовое зрение по специально отобраным таблицам (например, Ишихары).

Необходимое количество оценщиков зависит от чувствительности и надежности принятой процедуры испытаний и от ожидаемого размера искомого эффекта.

Для облегчения дальнейшего изучения этого фактора экспериментаторам следует включать как можно больше подробностей о характеристиках своих экспертных групп. В частности, предлагается указать род занятий (например, сотрудник телерадиовещательной организации, студент университета, офисный работник), пол и возраст.

А6-2.6 Проектирование эксперимента

Экспериментатор планирует эксперимент в соответствии с конкретными требованиями по затратам и точности. Рекомендуется включить в эксперимент по крайней мере две реплики (то есть два повтора идентичных условий). Реплики позволяют рассчитать индивидуальную достоверность и при необходимости отбросить ненадежные результаты, полученные от отдельных участников. К тому же реплики в какой-то мере обеспечивают уравнивание эффектов обучения в ходе испытаний. Дальнейшее улучшение компенсации эффектов обучения достигается путем включения в испытания в начале каждого сеанса нескольких "тренировочных демонстраций". Эти условия должны быть репрезентативными в отношении последующих демонстраций сеанса. Предварительные демонстрации не принимаются во внимание при статистическом анализе результатов испытаний.

Сеанс, то есть серия демонстраций, должен длиться не более получаса.

При испытании нескольких сцен или алгоритмов они должны демонстрироваться в случайном порядке. Случайный порядок можно изменить таким образом, чтобы одни и те же сцены или алгоритмы не демонстрировались непосредственно друг за другом (то есть последовательно).

А6-3 Методы оценки

Видеохарактеристики мультимедийных систем можно исследовать с использованием методик, описанных в части 2. Для их оценки можно применять метод субъективной оценки качества мультимедийного видеоизображения (SAMVIQ), в котором используются преимущества характеристик мультимедийных программ.

Приложение 7 к части 3

Методы субъективной оценки систем стереоскопического 3D-телевидения

A7-1 Показатели (воспринимаемые) оценки

Стереоскопическое 3D-телевидение использует характеристики бинокулярной зрительной системы человека, воссоздавая условия, приводящие к восприятию относительной глубины объектов визуальной обстановки. В настоящее время основное требование к созданию стереоскопических изображений – наличие по меньшей мере двух видов одного места с двух находящихся на одной горизонтальной линии камер. Изображения присутствующих объектов будут иметь разные относительные положения на левостороннем и правостороннем изображениях. Это различие относительных положений в двух изображениях обычно называется расхождением изображений (или параллаксом) и, как правило, выражается в пикселях, физических величинах (например, мм) или относительных величинах (например, процентах ширины экрана). Расхождение изображений следует отличать от углового (относящегося к сетчатке глаза) расхождения. Фактически одна и та же информация о расхождении изображений даст различные угловые (относящиеся к сетчатке глаза) расхождения при различных расстояниях просмотра. Масштаб и направление восприятия глубины основывается на масштабе и направлении относящихся к сетчатке расхождений, вызываемых стереоскопическим изображением.

Факторы оценки, обычно применяемые к моноскопическим телевизионным изображениям, такие как разрешение, передача цвета, отображение движения, общее качество, резкость и т. д., можно применять и к стереоскопическим телевизионным системам. Наряду с этим многие факторы будут относиться только к стереоскопическим телевизионным системам. Это могут быть такие факторы, как разрешение по глубине, которое представляет собой пространственное разрешение в направлении глубины, движение по глубине, то есть равномерно ли передается движение в направлении глубины, и пространственные искажения. Два хорошо известных примера последних: *эффект кукольного театра*, когда объекты воспринимаются как неестественно большие или маленькие, и *эффект картона*, когда объекты воспринимаются стереоскопически, но выглядят неестественно тонкими.

Можно определить три основных воспринимаемых измерения, которые в совокупности влияют на оценку пользователем качества услуги, предоставляемой стереоскопической системой: *качество изображения, качество глубины и зрительный комфорт*. Некоторые исследователи утверждают, что психологическое воздействие технологий стереоскопической передачи изображений также можно измерить относительно более общих понятий, таких как *естественность* и *ощущение присутствия*.

A7-1.1 Первичные воспринимаемые показатели

Под *качеством изображения* понимается воспринимаемое качество изображения, предоставляемого системой. Это один из основных определяющих показателей качества видеосистемы. На качество изображения в основном влияют технические параметры и ошибки, привносимые, например, процессами кодирования и/или передачи.

Под *качеством глубины* понимается способность системы передавать усовершенствованное ощущение глубины. Наличие монокулярных ориентиров, таких как линейная перспектива, размытие изображения, градиенты и т. д., создает определенное ощущение глубины даже в стандартных 2D изображениях. Вместе с тем стереоскопические 3D-изображения также содержат информацию о расхождении изображений, которая дает дополнительную информацию по глубине и тем самым обостренное ощущение глубины по сравнению с 2D.

Под *зрительным (дис)комфортом* понимается субъективное ощущение (дис)комфорта, которое может быть связано с просмотром стереоскопических изображений. Недолжным образом заснятые или недолжным образом показываемые стереоскопические изображения могут быть серьезным источником дискомфорта.

A7-1.2 Дополнительные воспринимаемые показатели

Под *естественностью* понимается восприятие стереоскопического изображения как верного отображения действительности (то есть реализм восприятия). В стереоскопическом изображении могут присутствовать различные типы искажений, делающие его менее естественным. Так, стереоскопические объекты иногда воспринимаются как неестественно большие или маленькие (эффект кукольного театра), или же они кажутся неестественно тонкими (эффект картона).

Под *ощущением присутствия* понимается субъективное ощущение присутствия в одном месте или одной обстановке при фактическом нахождении в другом месте или другой обстановке.

В настоящей Рекомендации представлена информация относительно методов и процедур оценки трех первичных показателей: качества изображения, качества глубины и зрительного комфорта, кратко описанных выше. Методики оценки естественности и ощущения присутствия не включены в настоящую Рекомендацию, но их планируется включить позже.

A7-2 Методики субъективной оценки

В настоящей Рекомендации описываются различные методики оценки качества изображения. Во всех методах набор последовательностей видеокадров, обработанный исследуемыми системами (например, алгоритм с различными параметрами; технология кодирования с различными скоростями передачи; различные сценарии передачи и т. д.), показывается группе зрителей в ходе ряда оценочных испытаний. В ходе каждого испытания зрителям предлагается оценить соответствующую характеристику (например, качество изображения) последовательности(ей) видеокадров по предписываемой шкале. Различные методы отличаются друг от друга в основном в отношении режима представления, то есть способа, которым видеокадр(ы) представляются зрителям, и шкалы, используемой зрителями для ранжирования этих последовательностей кадров.

Испытательные изображения представляют собой бинокулярные стереоизображения, отобранные на основе критериев, перечисленных в разделе A7-4. Оценивающие дают оценку следующим трем показателям:

- качество изображения: воздействие на разрешение стереоскопических 3D-изображений системы, в которой имеется путь между испытательными изображениями и дисплеем, используемым для показа изображений, которые предстоит оценить;
- качество глубины: воздействие на восприятие глубины стереоскопических 3D-изображений системы, в которой имеется путь между испытательными изображениями и дисплеем, используемым для показа изображений, которые предстоит оценить;
- зрительный комфорт: воздействие на удобство просмотра стереоскопических 3D-изображений системы, в которой имеется путь между испытательными изображениями и дисплеем, используемым для показа изображений, которые предстоит оценить.

В настоящее Приложение включены шесть методов из настоящей Рекомендации; эти методы успешно применяются на протяжении последних двадцати лет для решения исследовательских вопросов, связанных с качеством изображения, качеством глубины и зрительным комфортом технологий передачи стереоскопических изображений. Это следующие методы:

- метод с одним источником воздействия (SS);
- метод с двумя источниками воздействия и с использованием шкалы искажений (DSIS);
- метод двух источников воздействия с непрерывной шкалой качества (DSCQS);
- метод сравнения воздействий (SC);
- метод непрерывной оценки качества при одном источнике воздействия (SSCQE);
- метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE).

Когда это было целесообразно, методы использовались в слегка видоизмененной форме, например различные шкалы для зрительного комфорта. Режим представления и шкалы, связанные с методом оценки качества изображения, качества глубины и зрительного комфорта, представлены соответственно в таблицах 3-10, 3-11 и 3-12.

В настоящем разделе далее дается краткое описание каждой из методик. В последующих разделах представлены методические элементы, общие для всех методов.

ТАБЛИЦА 3-10

Субъективный метод оценки качества изображения





Способ представления	Продолжительность последовательности	Двоичная шкала	Дискретная шкала	Непрерывная шкала
Методы с одним источником воздействия (SS), описанные в Приложении 3 к части 2	~10 с		5 Отлично 4 Хорошо 3 Удовлетворительно 2 Плохо 1 Неприемлемо	
Метод с двумя источниками воздействия с использованием шкалы искажений (DSIS), описанный в Приложении 1 к части 2, пп. А1-1–А1-5			5 Незаметно 4 Заметно, но не раздражает 3 Слегка раздражает 2 Раздражает 1 Очень раздражает	
Метод двух источников воздействия с непрерывной шкалой качества (DSCQS), описанный в Приложении 2 к части 2, пп. А2-1–А2-3	~10 с			
Методы сравнения воздействий (SC), описанные в Приложении 4 к части 2	~10 с	А в сравнении с В	-3 Намного хуже -2 Хуже -1 Несколько хуже 0 То же 1 Несколько лучше 2 Лучше 3 Намного лучше	
Метод непрерывной оценки качества при одном источнике воздействия (SSCQE), описанный в Приложении 5 к части 2	~3–5 мин.			
Метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE), описанный в Приложении 6 к части 2, пп. А6-1–А6-4				

ТАБЛИЦА 3-11

Субъективный метод оценки качества глубины









Способ представления	Продолжительность последовательности	Двоичная шкала	Дискретная шкала	Непрерывная шкала
Методы с одним источником воздействия (SS), описанные в Приложении 3 к части 2	~10 с		5 Отлично 4 Хорошо 3 Удовлетворительно 2 Плохо 1 Неприемлемо	 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Метод с двумя источниками воздействия с использованием шкалы искажений (DSIS), описанный в Приложении 1 к части 2, пп. А1-1–А1-5			5 Незаметно 4 Заметно, но не раздражает 3 Слегка раздражает 2 Раздражает 1 Очень раздражает	
Метод двух источников воздействия с непрерывной шкалой качества (DSCQS), описанный в Приложении 2 к части 2, пп. А2-1–А2-3	~10 с			 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Методы сравнения воздействий (SC), описанные в Приложении 4 к части 2	~10 с	А в сравнении с В	-3 Намного хуже -2 Хуже -1 Несколько хуже 0 То же 1 Несколько лучше 2 Лучше 3 Намного лучше	
Метод непрерывной оценки качества при одном источнике воздействия (SSCQE), описанный в Приложении 5 к части 2	~3–5 мин.			 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE), описанный в Приложении 6 к части 2, пп. А6-1–А6-4				 Идеальная четкость (код 100) Нулевая четкость (код 0)

ТАБЛИЦА 3-12

Субъективный метод оценки зрительного комфорта

Способ представления	Продолжительность последовательности	Двоичная шкала	Дискретная шкала	Непрерывная шкала
Методы с одним источником воздействия (SS), описанные в Приложении 3 к части 2	~10 с		5 Отлично 4 Хорошо 3 Удовлетворительно 2 Плохо 1 Неприемлемо	 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Метод с двумя источниками воздействия с использованием шкалы искажений (DSIS), описанный в Приложении 1 к части 2, пп. А1-1–А1-5			5 Незаметно 4 Заметно, но не раздражает 3 Слегка раздражает 2 Раздражает 1 Очень раздражает	
Метод двух источников воздействия с непрерывной шкалой качества (DSCQS), описанный в Приложении 2 к части 2, пп. А2-1–А2-3	~10 с			 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Методы сравнения воздействий (SC), описанные в Приложении 4 к части 2	~10 с	А в сравнении с В	-3 Намного хуже -2 Хуже -1 Несколько хуже 0 То же 1 Несколько лучше 2 Лучше 3 Намного лучше	
Метод непрерывной оценки качества при одном источнике воздействия (SSCQE), описанный в Приложении 5 к части 2	~3–5 мин.			 Отлично Хорошо Удовлетворительно Плохо Неприемлемо
Метод непрерывной оценки с двумя источниками одновременного воздействия (SDSCE), описанный в Приложении 6 к части 2, пп. А6-1–А6-4				 Идеальная четкость (код 100) Нулевая четкость (код 0)

А7-3 Общие условия просмотра

Условия просмотра (включая яркость экрана, контрастность, фоновую освещенность, расстояние просмотра и т. д.) должны соответствовать условиям просмотра, используемым для изображения 2D, как это описано в разделе 2.1 части 1. Имеются два обоснования для такого согласованного подхода. Во-первых, на практике пользователи будут смотреть 3D-телевидение на тех же самых дисплеях и при тех же условиях просмотра, что и при 2D ТВ. Во-вторых, потребуется часто измерять прогресс в качественных характеристиках видеотехнологий 3D-ТВ по отношению (то есть "в сравнении") к прогрессу в стандартных видеотехнологиях ТВЧ.

В разделе 2.1 части 1 указаны два возможных критерия для выбора расстояния просмотра. Следует выбирать расчетное расстояние просмотра (DVD). Для цифровой системы DVD является расстоянием, на котором два соседних пикселя составляют в глазу зрителя угол, равный 1 угловой минуте.

Следует отметить, что, поскольку два соседних пикселя составляют в глазу зрителя угол, равный 1 угловой минуте, на расчетном расстоянии просмотра наименьшее угловое (от сетчатки глаза) расхождение, которое может быть представлено системой (то есть разрешение системы по глубине), равно 1 угловой минуте (или, что то же самое, 60 угловым секундам). Результаты исследования показали, что примерно 97% населения могут различать расхождения по горизонтали, равные или меньшие 140 угловых секунд, и по меньшей мере 80% населения могут обнаруживать расхождения по горизонтали в 30 угловых секунд. Таким образом, большинство наблюдателей не будут сталкиваться ни с какими трудностями с анализом незначительных расхождений, имеющихся в современных видеосистемах 3D, на расчетном расстоянии просмотра.

A7-4 Материал для испытаний

Выбор материала для испытаний должен обосновываться прикладным вопросом, изучаемым в рамках исследования. В целом контент испытательных последовательностей (спорт, постановочные программы, фильмы и т. д.) и их пространственно-временные характеристики должны быть репрезентативными для программ, передаваемых исследуемой службой.

Кроме того, избранный контент стереоскопических испытательных последовательностей должен быть, как правило, удобным для просмотра. Зрительный комфорт стереоскопических изображений зависит в значительной степени от расхождений в изображениях (параллакс), имеющихся в изображениях и условиях просмотра. В связи с этим следует принимать меры для обеспечения того, чтобы расхождения не превышали пределов, указанных в приведенных ниже разделах, если только исследование особо не направлено на измерение зрительного комфорта. Кроме того, следует измерять и сообщать, когда это возможно, статистические данные: среднее, стандартное отклонение и диапазон (миним./максим.) распределения расхождений в испытательной последовательности.

Параллакс – несоответствия между левым и правым изображениями, и распределение и изменение параллакса можно предложить в качестве вопросов, которые следует учитывать при выборе испытываемых изображений как удобных для просмотра стереоскопических 3D-изображений. Отношения между удобным для просмотра стереоскопическим 3D-изображением и параллаксом, несоответствие между левосторонним и правосторонним изображениями, а также распределение и изменение параллакса описываются в последующих подразделах.

A7-4.1 Использование эталонного видеоматериала

Исследователи могут пожелать включить, при наличии, эталонные последовательности в качестве части набора испытательных последовательностей. Эти последовательности, как правило, являются такой версией испытательных последовательностей, которые не подвергались никакой обработке (то есть это исходные последовательности источника). Для стереоскопических исследований главным эталоном является исходная необработанная стереоскопическая последовательность. Однако план эксперимента может также включать моноскопическую эталонную версию (то есть только одно изображение исходной последовательности источника); например, при исследованиях зрительного комфорта может оказаться полезным использовать в качестве отправной точки зрительный комфорт моноскопического эталона. Эту моноскопическую версию эталона следует представить в режиме 3D (например, левое изображение представляется и для левого, и для правого глаза с использованием одних и тех же установок аппаратного обеспечения 3D как для фактической стереоскопической последовательности). Включение этого эталона в план эксперимента дает два важных преимущества. Во-первых, оно дает возможность измерить прозрачность (известную и как точность воспроизведения), которую обеспечивает изучаемый алгоритм или технология⁸.

⁸ Прозрачность (точность воспроизведения) – это концепция, описывающая качество работы кодека или системы по отношению к идеальной системе передачи без каких-либо ухудшений. Легко видеть, что прозрачность может быть измерена путем сравнения показателей, присвоенных эталонной последовательности, с показателями, присвоенными последовательности, обработанной исследуемым алгоритмом или технологией.

Во-вторых, включение эталона обеспечивает точки привязки к высокому качеству, что может помочь в стабилизации показателей⁹.

A7-4.2 Пределы визуального комфорта

Чрезмерное расхождение, или параллакс, вызывает визуальный дискомфорт, возможно, потому что оно усиливает конфликт между аккомодацией и вергентностью. Поэтому было предложено с целью минимизации конфликта аккомодации-вергентности, чтобы расхождение стереоскопического изображения было достаточно малым, с тем чтобы воспринимаемая глубина объектов попадала в пределы "зоны комфорта". Для определения этих пределов предложено несколько подходов. В одном подходе используется мера экранного параллакса, выраженная в процентах от горизонтального размера экрана, для того чтобы указать пределы комфортного просмотра. Была предложена величина пересекающегося/отрицательного расхождения, равная 1%, и величина непересекающегося/положительного расхождения, равная 2% (суммарная величина около 3%). Согласно другому подходу, комфортная зона разграничивается глубиной резкости глаза. Для условий просмотра, типичных для телевизионного вещания, исследователи приняли глубину резкости в пределах от $\pm 0,2D$ (диоптрий) до $\pm 0,3D$ (диоптрий). Для систем с разрешением изображения ТВЧ 1920×1080 (Рек. МСЭ-R ВТ.709), наблюдаемых с проектного расстояния просмотра, равного $3,1H$, эти величины приблизительно соответствуют $\pm 2\%$ и $\pm 3\%$ экранного параллакса. И наконец, в третьем подходе пределы комфортности указываются в значениях расхождения на сетчатке, и эти пределы устанавливаются равными $\pm 1^\circ$ от угла зрения как для положительного, так и для отрицательного расхождения.

Причем эти различные подходы стремятся к одним и тем же пределам комфортности. Напомним, что на расчетном расстоянии просмотра два соседних пикселя составляют угол, равный 1 угловой минуте, в глазу зрителя. Таким образом, 60 пикселей соответствуют углу зрения 1° . Это позволяет нам легко указать пределы комфортности в значениях расхождения на сетчатке (для усредненного зрителя). Например, для систем с разрешением изображения ТВЧ 1920×1080 (Рекомендация МСЭ-R ВТ.709) 1% (~19,2 пикселя) соответствует приблизительно 20 угловым минутам, 2% – ~40 угловым минутам и 3% – ~60 угловым минутам (что эквивалентно 1°).

Следует отметить, что даже хотя на расчетном расстоянии просмотра два соседних пикселя всегда образуют угол, равный 1 угловой минуте, физический разнос (то есть выраженный в мм) этих пикселей с увеличением размера экрана возрастает (количество пикселей остается тем же самым, а физический размер экрана возрастает). Следовательно, более высокие пределы (например, $\pm 3\%$) могут привести к большему размеру экрана на некотором физическом расстоянии между соответствующими точками (то есть параллаксу двух изображений в мм), который превышает расстояние между зрачками усредненного зрителя (~63–65 мм). Это может привести к увеличению дискомфорта.

A7-4.3 Несоответствия между левым и правым изображениями

В стереосистемах 3D, бинокулярное 3D-изображение формируется представлением левого и правого изображения в соответствующий глаз. Если между этими двумя изображениями возникнут несоответствия, то они могут вызвать психофизический стресс, а в некоторых случаях может нарушиться трехмерность просмотра. Например, когда снимаются и воспроизводятся стереоскопические программы 3D-ТВ, могут возникать геометрические искажения, такие как несоответствие размеров, вертикальное смещение и ошибка поворота между левым и правым изображениями. Было бы предпочтительно, чтобы испытательные изображения были свободны от этих геометрических искажений. Для получения дополнительной информации см. пункт 3.2.1 Приложения 4 к Отчету МСЭ-R ВТ.2160-2.

При выборе испытательных изображений в качестве удобных для просмотра стереоскопических 3D-изображений следует учитывать следующие факторы, касающиеся несоответствий между левым и правым изображениями:

⁹ Считается, что стабильность показателей по пространству (т. е. по разным лабораториям) и времени (т. е. в одной и той же лаборатории, но в разное время) может быть улучшена также использованием точек привязки к низкому качеству. Однако в МСЭ отсутствуют планы по скорейшему созданию/определению стандартизованных точек привязки к низкому качеству для оценки технологий стереоскопического представления изображений.

- геометрическое несоответствие, включая размер, вертикальное смещение и вращение;
- несоответствие яркости, включая уровень белого и черного;
- перекрестные помехи.

A7-4.4 Диапазон, распределение и изменение параллакса

При стереоскопических изображениях распределение параллакса коррелируется с визуальным комфортом.

Распределение параллакса стереоскопических изображений не является непрерывным при смене сюжетов в кадрах. Случаи значительного параллакса или резкие изменения параллакса создают дискомфорт при просмотре, поэтому важно тщательно распределять параллакс в испытательных изображениях. Для получения дополнительной информации см. пункт 3.2.2 Приложения 4 к Отчету МСЭ-R ВТ.2160-2.

В общем случае, поскольку исследования с использованием стереоскопических испытательных последовательностей могут вызывать некоторую степень дискомфорта просмотра, рекомендуется использовать, когда это возможно, материалы для испытаний, в которых расхождение не превышает пределов комфорта, хотя случайные отклонения выше этих пределов комфорта могут допускаться.

A7-5 Экспериментальная аппаратура

Экспериментальная аппаратура (сервер видеосигналов, экран и т. д.) должна быть способна отображать испытательные последовательности с полным HD разрешением, например, используя формат упаковки кадров HDMI. Это даст большую гибкость выбора спектра исследований, которые могут быть выполнены.

В настоящее время отсутствуют стандартизованные эталонные экраны для оценки 3D-ТВ. Соответственно ожидается, что большинство исследователей будут использовать существующие экраны 3D-ТВ потребительского уровня. Поскольку характеристики таких экранов могут варьироваться в зависимости от производителей, исследователи настоятельно призываются сообщать соответствующую информацию о настройках экрана, используемого в этих исследованиях.

A7-6 Наблюдатели

A7-6.1 Размер выборки наблюдателей

В общем случае рекомендуется использовать по меньшей мере 30 зрителей. Однако признается, что фактическое количество будет зависеть от конкретных задач исследования, отмечая, что соображения по выбору размера выборки в исследованиях, касающихся 3D, такие же, как и в исследованиях, касающихся 2D.

A7-6.2 Отсевание по зрению

Наблюдателей следует отсеивать по остроте зрения, дальтонизму, стереоскопичности зрения используя существующие клинические испытания зрения, такие как таблицы Снеллена, эквивалентные для остроты зрения; таблицы Ишихары или эквивалентные им для цвета; и Рандо или эквивалентные им для стереоскопичности зрения. Отметим, что в испытаниях на стереоскопичность зрения, подобных испытаниям Рандо, Stereo Fly или Фрисби обычно измеряется расхождение приблизительно от 20 до 400 угловых секунд. Исследователи призываются сообщать соответствующую статистику о стереоскопических способностях наблюдателей, участвующих в исследовании. Если требуется более детальный анализ стереоскопических способностей участников, исследователи могут использовать материалы для испытаний, показанные в Прилагаемом документе 1 к настоящему Приложению.

A7-7 Инструкции для наблюдателей

Инструкции следует индивидуализировать для конкретных изучаемых величин (например, качества глубины зрения, комфорта и т. д.). При этом этические указания для исследований трехмерности жестче, чем типовые указания, используемые при оценке качества 2D изображений, поскольку

участники могут испытывать зрительный дискомфорт. В общем случае при исследованиях трехмерности требуется уделять больше внимания информированию участника о мотивации данного исследования, а также о всех возможных негативных эффектах, возникающих в результате стимулирующих воздействий, используемых при данном исследовании.

A7-8 Продолжительность сеанса

Если просматриваемый материал считается комфортным, то продолжительность испытательного сеанса может быть такой же, какая используется при 2D исследованиях (то есть ~20–40 минут, перемежаемые перерывами). Если известно, что материал содержит чрезмерный параллакс, и значит, что он потенциально некомфортен, то продолжительность сессии следует ограничить.

A7-9 Изменчивость ответов

Показатели, сообщаемые зрителями во время экспериментов по субъективной оценке, в общем случае довольно изменчивы. Различия в показателях зрителей могут просто отражать характеристики отобранных для сравнения зрителей, и они могут быть устранены увеличением размера выборки.

Однако часть изменчивости может быть отнесена на изменения структуры ответов индивидуальных пользователей во время эксперимента. Эти изменения предполагают изменение критериев оценки, которое может возникать из-за приобретения опыта в ходе решения задачи, изучения характеристик артефактов и т. п. Для того чтобы минимизировать негативное влияние такой изменчивости, исследователям следует разработать соответствующие процедуры профессиональной подготовки (выполнение задачи, уровень ухудшения и т. д.), использовать многократную рандомизацию (то есть представлять разным зрителям испытательные последовательности в различном случайном порядке), а также использовать дублирование (которое позволило бы измерить возможные изменения в структуре ответов).

A7-10 Критерии исключения зрителей

Критерии исключения зрителей (отбор наблюдателей) для методов, указанных в разделе A7-2, описаны в части 1.

A7-11 Статистический анализ

Методы статистического анализа при изучении систем трехмерного представления изображений такие же, как и для систем двумерного представления изображений.

Прилагаемый документ 1 к Приложению 7

Испытательные материалы для испытаний зрения

A7-1 Испытания зрения

В таблице 3-13 перечислены испытательные таблицы для испытаний зрения. Эти 12 испытаний отобраны в соответствии с иерархией системы человеческого зрения от нижнего уровня до высшего уровня. Ниже описаны восемь основных испытаний зрения (VT), остальные четыре – клинические испытания. Наблюдатели должны иметь нормальное бинокулярное зрение, означающее, что они должны успешно пройти испытание VT-04 по точности бинокулярного зрения и VT-07 по динамике бинокулярного зрения. Оставшиеся шесть испытаний предназначены для получения более детальных характеристик. Испытательные таблицы следует просматривать с расстояния, равного утроенной высоте экрана дисплея.

ТАБЛИЦА 3-13

Стереоскопические испытательные материалы для испытаний зрения

№	Испытание	Цель испытания	Содержание
1	Одновременное восприятие	Способность одновременного и в правильном положении восприятия дихоптически представленных изображений	Для одного глаза представлено изображение клетки, а для другого – льва
2	Бинокулярное слияние	Способность воспринимать два дихоптических изображения левым и правым глазами как одно изображение	Изображение для одного глаза имеет две, а для другого – три точки, причем одна точка является общей
3	Стереоскопическое восприятие крупных деталей	Способность воспринимать дихоптически представленные изображения с параллаксом как одно изображение с глубиной крупных деталей	Изображения для двух глаз являются стереопарой стрекозы с расправленными крыльями
4	Стереоскопическое восприятие мелких деталей	Способность воспринимать дихоптически представленные изображения с параллаксом как одно изображение с глубиной мелких деталей	Представлены девять небольших фигурок ромбовидной формы, каждая из них имеет четыре круга, в которых только один круг имеет малый параллакс
5	Предел слияния с пересечением	Способность воспринимать дихоптически представленные изображения с пересекающимися расхождениями как одно изображение	Стереопара вертикальных черточек представлена со своим пересекающимся параллаксом, изменяющимся со скоростью 10'/с
6	Предел слияния без пересечения	Способность воспринимать дихоптически представленные изображения с непересекающимися расхождениями как одно изображение	Стереопара вертикальных черточек представлена со своим непересекающимся параллаксом, изменяющимся со скоростью 11'/с
7	Динамическое стереоскопическое восприятие	Способность воспринимать глубину в движущихся изображениях, образованных стереограммами произвольно выбранных точек	Динамическая стереограмма из произвольно выбранных точек
8	Острота бинокулярного зрения	Острота бинокулярного зрения с бинокулярным слиянием, включая любое нарушение остроты монокулярного зрения, которое может помешать хорошему стереоскопическому восприятию	Буквы E разнообразной ориентации и размера
9	Горизонтальное косоглазие	Горизонтальная девиация глаза, с которой пациент не может справиться	Вертикальные и горизонтальные линии
10	Вертикальное косоглазие	Вертикальная девиация глаза, с которой пациент не может справиться	Вертикальные и горизонтальные линии
11	Анизейкония	Состояние, при котором зримое изображение объекта, наблюдаемого одним глазом, отличается по размеру и форме от изображения, наблюдаемого другим глазом	Левое изображение состоит из знаков "[o", а правое – из знаков "o]", причем расположение буквы "o" является общим
12	Циклофория	Девиация одного или другого глаза от переднезадней оси, когда слияние предотвращено	Левое изображение состоит из циферблата, а правая – из часовых стрелок, указывающих на шесть часов

ПРИМЕЧАНИЕ 1. – Эти материалы представлены в формате 1125/60/I (см. Рекомендацию МСЭ-R ВТ.709).

ПРИМЕЧАНИЕ 2. – Материалы могут быть получены из Института инженеров по информации изображений и телевидению (ITE), 3-5-8 Шибакен, Минато-ку, Токио 105-0011, Япония, тел.: 81-3-3432-4675, электронная почта: ite@ite.or.jp.

Внизу: правое и левое изображения расположены непосредственно рядом с целью объяснения слияния без пересечения.

1) ВТ-01: Одновременное восприятие (испытание со львом)

Проверяется способность одновременного и в правильном положении восприятия дихоптически представленных изображений. Для одного глаза представлено изображение клетки, а для другого – льва, причем его положение меняется со скоростью 12'/с. Размер каждого изображения установлен в 10°, что позволяет наблюдателям фиксировать изображения в своей парамакулярной зоне. Наблюдатели с нормальным зрением могут видеть льва в клетке в определенное время периода представления.

РИСУНОК 3-8

Испытательная таблица для ВТ-01



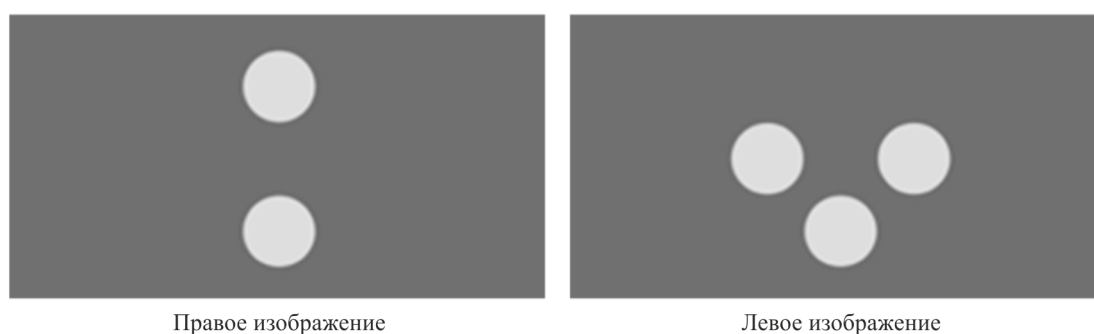
ВТ.0500-03-8

2) ВТ-02: Бинокулярное слияние (метод Уорга)

Проверяется способность воспринимать два дихоптические изображения левым и правым глазами как одно изображение. Изображение для одного глаза имеет две, а для другого – три точки, причем одна точка является общей. Наблюдатели с нормальным зрением могут видеть четыре точки.

РИСУНОК 3-9

Испытательная таблица для ВТ-02



ВТ.0500-03-9

3) ВТ-03: Стереоскопическое восприятие крупных деталей (испытание со стрекозой)

Проверяется способность воспринимать дихоптически представленные изображения с параллаксом как одно изображение с глубиной крупных деталей. Изображения для двух глаз являются стереопарой стрекозы с расправленными крыльями. Наблюдатели с нормальным зрением могут воспринимать крылья спереди экрана.

РИСУНОК 3-10

Испытательная таблица для ВТ-03



Правое изображение

Левое изображение

ВТ.0500-03-10

4) ВТ-04: Стереоскопическое восприятие мелких деталей (испытание с кругами)

Проверяется способность воспринимать дихоптически представленные изображения с параллаксом как одно изображение с глубиной мелких деталей. Ниже на рисунке представлены девять небольших фигурок ромбовидной формы для испытания, причем каждая из них имеет четыре круга, в которых только один круг имеет малый параллакс. Наблюдатели с нормальным зрением могут воспринимать круг с малым параллаксом спереди экрана. В таблице 3-14 показаны номер испытания, правильные ответы и угол стереоскопического восприятия на расстоянии $3H$.

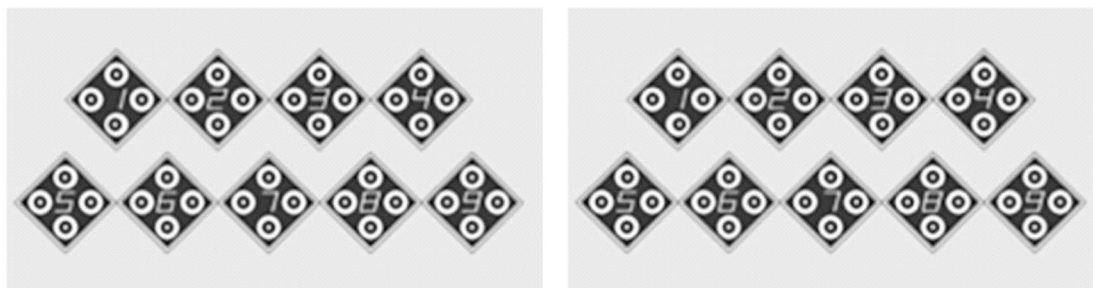
ТАБЛИЦА 3-14

Правильные ответы и параллакс

№ испытания	Правильные ответы	Угол стереоскопического восприятия на расстоянии $3H$ (")
1	Внизу	480
2	Слева	420
3	Внизу	360
4	Вверху	300
5	Верху	240
6	Слева	180
7	Справа	120
8	Слева	60
9	–	0

РИСУНОК 3-11

Испытательная таблица для VT-04



Правое изображение

Левое изображение

ВТ.0500-03-11

5) VT-05: Предел слияния с пересечением (испытание с вертикальной чертой)

Проверяется способность воспринимать дихоптически представленные изображения с пересекающимися расхождениями как одно изображение. Стереопара вертикальных черточек представлена со своим параллаксом, изменяющимся со скоростью 10°/с. Пределы слияния для возрастающего и убывающего рядов можно измерить. Наблюдателям предлагается сообщать о перерыве слияния, как только они заметят сдвоенное изображение в возрастающем ряде, и о восстановлении слияния, как только они заметят дихоптические изображения как одно изображение в убывающем ряде.

РИСУНОК 3-12

Испытательная таблица для VT-05



Правое изображение

Левое изображение

ВТ.0500-03-12

6) VT-06: Предел слияния без пересечения (испытание с вертикальной чертой)

Проверяется способность воспринимать дихоптически представленные изображения с непересекающимися несовпадениями как одно изображение. Представленные изображения – те же, что и в случае с пересечением, рассмотренным выше, однако правое и левое изображения поменялись местами.

РИСУНОК 3-13

Испытательная таблица для VT-06



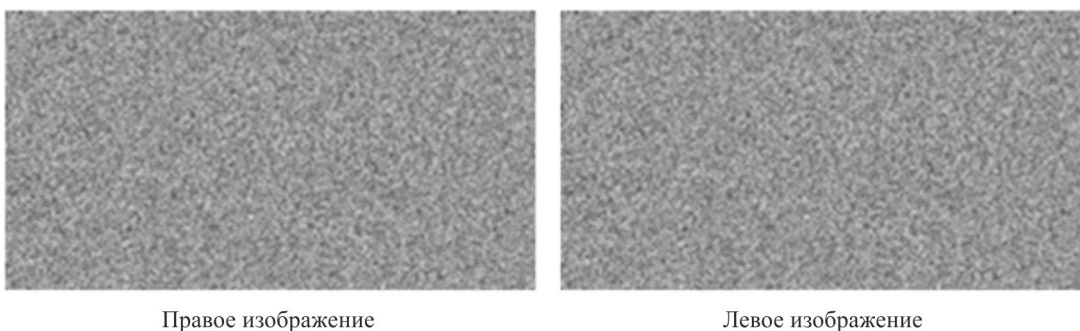
ВТ.0500-03-13

7) VT-07: Динамическое стереоскопическое восприятие (испытание с динамической стереограммой произвольно выбранных точек)

Проверяется способность воспринимать глубину в движущихся изображениях, образованных стереограммами произвольно выбранных точек. Наблюдатели с нормальным зрением могут воспринимать прямоугольную форму и синусоидальное движение глубины в динамической стереограмме произвольно выбранных точек.

РИСУНОК 3-14

Испытательная таблица для VT-07



ВТ.0500-03-14

8) VT-08: Острота бинокулярного зрения (испытание на остроту)

Проверяется острота бинокулярного зрения с бинокулярным слиянием, включая любое нарушение остроты монокулярного зрения, которое может помешать хорошему стереоскопическому восприятию. На изображениях видны четыре колонки и пять линий, состоящие из букв Е с различным расположением и размерами. Две центральные колонки можно видеть обоими глазами, две левые колонки видны только левым глазом, а две правые колонки видны только правым глазом. Наблюдатели с нормальным зрением могут правильно определить расположение букв Е. Размеры букв соответствуют остроте в 1,0; 0,5; 0,33; 0,25 и 0,125 на расстоянии 3 Н.

РИСУНОК 3-15

Испытательная таблица для VT-08



Правое изображение

Левое изображение

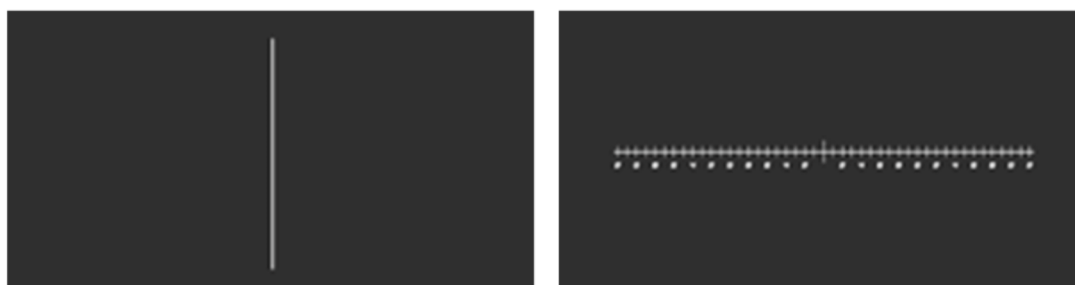
ВТ.0500-03-15

9 и 10) VT-09: Горизонтальное косоглазие (испытание горизонтального косоглазия методом Меддокса) и VT-10: вертикальное косоглазие (испытание вертикального косоглазия методом Меддокса)

Эти испытательные таблицы измеряют горизонтальную и вертикальную девиацию глаза. Визуальные оси допускают расположение относительно друг друга, отличное от расположения, требуемого физиологическими условиями. Изображения состоят из вертикальных и горизонтальных линий. Наблюдатели с нормальным зрением могут воспринимать точку пересечения линий, расположенную приблизительно в центре этих линий. Единица измерения значений у отсечек – есть призматический диоптрий с PD (межзрачковое расстояние) = 65 мм на расстоянии 3,02 *H*.

РИСУНОК 3-16

Испытательная таблица для VT-09



ВТ.0500-03-16

РИСУНОК 3-17

Испытательная таблица для VT-10



ВТ.0500-03-17

11) VT-11: Анизейкония ("["]" испытание с буквами)

Состояние, при котором зримое изображение предмета, наблюдаемого одним глазом, по размеру и форме отличается от изображения, наблюдаемого другим глазом. Левое изображение состоит из знаков "[о", а правое – из знаков "о]", причем расположение буквы "о" является общим. Наблюдатели с нормальным зрением могут воспринимать знаки "[" и "]" как имеющие одинаковый размер и одинаковую высоту.

РИСУНОК 3-18

Испытательная таблица для VT-11



ВТ.0500-03-18

12) VT-12: Циклофория (испытание с часами)

Девияция глаза от переднезадней оси только в том случае, если она охвачена, а слияние предотвращено. Левое изображение состоит из циферблата, а правая – из часовой стрелки, указывающей на шесть часов. Наблюдатели с нормальным зрением могут воспринимать часы только как шесть часов.

РИСУНОК 3-19

Испытательная таблица для VT-12



ВТ.0500-03-19