

Международный союз электросвязи

МСЭ-R

Сектор радиосвязи МСЭ

Рекомендация МСЭ-R ВТ.500-12
(09/2009)

**Методика субъективной оценки качества
телевизионных изображений**

Серия ВТ
Радиовещательная служба
(телевизионная)



Международный
союз
электросвязи

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 Приложении 1 к Резолюции 1 МСЭ-R. 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/en>, 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/publications/R-REC/en>.)

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	Подвижная спутниковая служба, спутниковая служба радиоопределения, любительская спутниковая 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 1 МСЭ-R.

Электронная публикация
Женева, 2010 г.

© ITU 2010

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.500-12

**Методика субъективной оценки
качества телевизионных изображений**

(Вопрос МСЭ-R 81/6)

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

Сфера применения

В настоящей Рекомендации содержится описание методик оценки качества изображения, включая общие методы испытания, шкалы оценок качества и условия просмотра. Рекомендуются метод шкалы искажений с двумя источниками воздействия (DSIS) и метод непрерывной шкалы качества с двойным источником воздействия, а также альтернативные методы оценки, такие как метод с одним источником воздействия (SS), методы сравнения воздействий, метод непрерывной оценки качества с одним источником воздействия (SSCQE) и метод непрерывной оценки с двумя источниками непрерывного воздействия (SDSCE).

Ассамблея радиосвязи МСЭ,

учитывая,

- a) что накоплено большое количество информации о методах оценки качества изображения, используемых в различных лабораториях;
- b) что изучение этих методов показывает, что между различными лабораториями наблюдается достаточно близкое совпадение по множеству аспектов таких испытаний;
- c) что для обмена информацией между различными лабораториями важно принять стандартизированные методы;
- d) что некоторые методы, 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 введение новых методов обработки телевизионного сигнала, таких как цифровое кодирование и уменьшение скорости цифрового потока, новых видов телевизионных сигналов, составляющие которых мультиплексированы с разделением по времени, и, возможно, новых служб, например, телевидения с повышенной четкостью и ТВЧ, может потребовать изменений в методах проведения субъективных оценок;
- f) что введение таких методов обработки, сигналов и служб увеличит вероятность того, что качество на каждом участке пути следования сигнала будет обусловлено процессами, происходящими в предыдущих частях этого пути,

рекомендует,

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

2 что в ближайшем будущем, несмотря на существование альтернативных методов и разработку новых методов, когда это возможно, должны использоваться методы, описанные в пп. 4 и 5 Приложения 1 к настоящей Рекомендации;

3 что, ввиду важности создания основы для субъективных оценок, во всех отчетах об испытаниях должно быть представлено самое полное, насколько возможно, описание конфигурации тестов, испытательных материалов, экспертов и методов;

4 что для того, чтобы облегчить обмен информацией между различными лабораториями, собранная информация должна быть статистически обработана по методике, подробно описанной в Приложении 2 к настоящей Рекомендации.

ПРИМЕЧАНИЕ 1. – Информация о методах субъективной оценки для определения качества телевизионных систем содержится в Приложении 1.

ПРИМЕЧАНИЕ 2. – Описание статистических методов обработки данных, собранных в ходе субъективных испытаний, содержится в Приложении 2.

Приложение 1

Описание методов оценки

1 Введение

Методы субъективных оценок 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 соответствующих субъективных испытаний, в первую очередь необходимо выбрать из различных имеющихся в наличии возможностей те, что соответствуют целям и условиям существующих проблем, для которых требуется оценка. Для облегчения решения данной задачи после описания общих особенностей, содержащегося в п. 2, в п. 3 приведена некоторая информация, касающаяся связанных с оценкой проблем, для которых пригоден каждый метод. Затем в пп. 4 и 5 подробно описаны два главных рекомендуемых метода. И, наконец, в п. 6 приведены общие сведения об альтернативных методах, исследуемых в настоящее время.

Цель данного Приложения ограничивается подробным описанием методов оценки. Выбор наиболее подходящего метода, тем не менее, зависит от задач обслуживания испытываемой системы. Поэтому полное описание процедур оценки для конкретных применений приводится в других Рекомендациях МСЭ-R.

2 Общие особенности

Представлены общие условия просмотра для проведения субъективных оценок. Конкретные условия просмотра в целях проведения субъективных оценок конкретных систем представлены в соответствующих рекомендациях.

2.1 Общие условия просмотра

Описаны разные среды с различными условиями просмотра.

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

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

Рассмотрены некоторые аспекты, касающиеся разрешения и контрастности мониторов.

2.1.1 Лабораторная среда

2.1.1.1 Общие условия просмотра для проведения субъективных оценок в лабораторной среде

Условия просмотра для экспертов должны быть следующими:

- | | | |
|----|---|--|
| a) | Отношение яркости бездействующего экрана к пиковой яркости: | $\leq 0,02$ |
| b) | Отношения яркости экрана при показе уровня черного в полностью затемненной комнате к показу уровня белого в тех же условиях: | $\approx 0,01$ |
| c) | Яркость и контрастность устройства отображения: | устанавливается при помощи генератора для настройки ВКУ (PLUGE) (см. Рекомендации МСЭ-R ВТ.814 и ВТ.815) |
| d) | Максимальный угол обзора по отношению к перпендикуляру (эта цифра относится к устройствам отображения с ЭЛТ, тогда как соответствующая цифра для других таких устройств является предметом изучения): | 30° |
| e) | Отношение яркости послесвечения к пиковой яркости изображения: | $\approx 0,15$ |
| f) | Цветность фона: | D_{65} |
| g) | Яркость других источников света в комнате просмотра: | низкая |

2.1.2 Домашняя среда

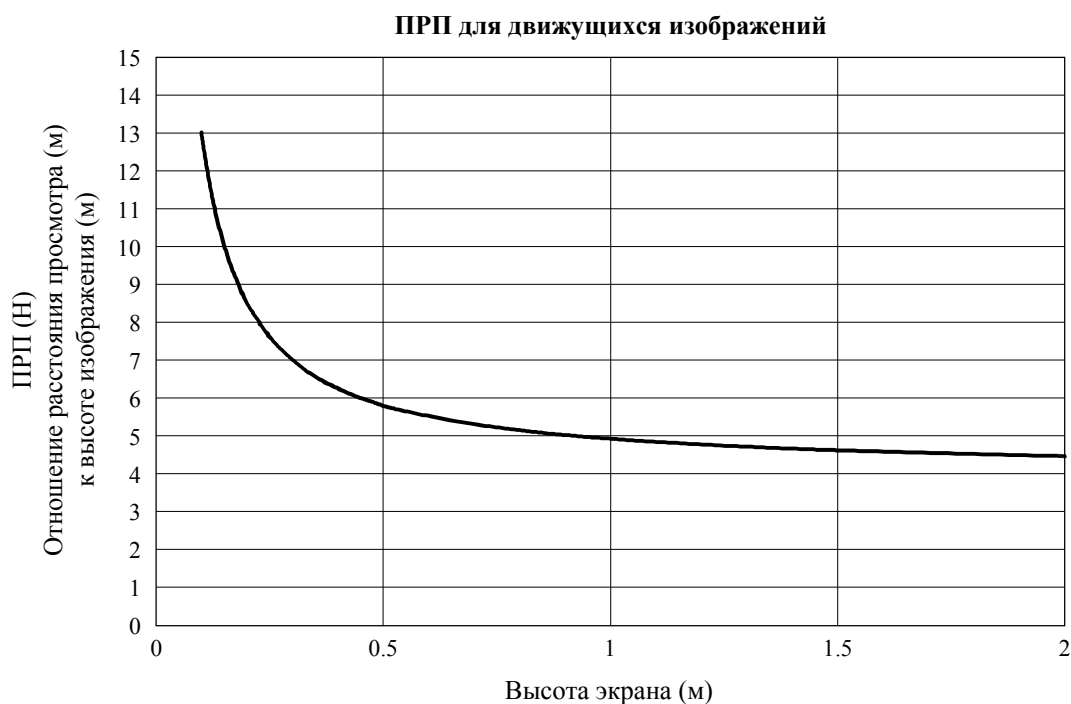
2.1.2.1 Общие условия просмотра для проведения субъективных оценок в домашней среде

- | | | |
|----|---|--|
| a) | Отношение яркости бездействующего экрана к пиковой яркости: | $\leq 0,02$ (см. п.2.1.4) |
| b) | яркость и контрастность устройства отображения: | устанавливается при помощи генератора для настройки ВКУ (PLUGE) (см. Рекомендации МСЭ-R ВТ.818 и ВТ.815) |
| c) | Максимальный угол обзора по отношению к перпендикуляру (эта цифра относится к устройствам отображения с ЭЛТ, тогда как соответствующая цифра для других таких устройств является предметом изучения): | 30° |
| d) | Размер экрана для формата изображения 4/3: | Этот размер экрана должен удовлетворять правилам предпочитаемого расстояния просмотра (ПРП) |
| e) | Размер экрана для формата изображения 16/9: | Этот экран должен удовлетворять правилам ПРП |

- | | | |
|----|--|------------------------|
| f) | Обработка в мониторе: | Без цифровой обработки |
| g) | Разрешение монитора: | см. п. 2.1.3 |
| h) | Пиковая яркость: | 200 кд/м ² |
| i) | Освещенность вокруг экрана (свет в среде, освещающий экран, должен измеряться перпендикулярно экрану): | 200 люкс |

Расстояние просмотра и размеры экрана должны выбираться с целью удовлетворения правилам ПРП. Правила ПРП (в зависимости от размеров экрана) представлены в таблице и на графике ниже. Цифры могут быть действительны для ТСЧ и ТВЧ, поскольку обнаруженное различие было очень небольшим.

Диагональ экрана (дюймы)		Высота экрана (Н) (м)	ПРП (Н)
формат 4/3	формат 16/9	(м)	(Н)
12	15	0,18	9
15	18	0,23	8
20	24	0,30	7
29	36	0,45	6
60	73	0,91	5
> 100	> 120	> 1,53	3–4



0500-00

Эти таблица и график предназначены, чтобы предоставить информацию о ПРП и соответствующих размерах экрана, и должны быть приняты в рекомендациях, касающихся конкретных применений.

2.1.3 Разрешение монитора

Разрешение профессиональных мониторов, оборудованных профессиональными ЭЛТ, обычно соответствует требуемым стандартам субъективной оценки в пределах их рабочего диапазона яркости.

Не все мониторы могут обеспечивать пиковую яркость в 200 кд/м².

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

Если для проведения субъективных оценок используется бытовой телевизор, разрешение может быть недостаточным в зависимости от значения яркости.

В этом случае 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 с целью проверки разрешения мониторов или бытовых телевизоров, является использование разворачиваемой испытательной таблицы, которая создается электронным способом.

Зрительный анализ позволяет проверить разрешение. Зрительный порог оценивается на уровне –12/–20 дБ. Основной недостаток этой системы заключается в эффекте наложения сигналов, создаваемого теневой маской, которая затрудняет зрительную оценку, однако, с другой стороны, наличие эффекта наложения указывает, что сигнал видеочастоты находится вне пределов, задаваемых теневой маской, следствием чего является недостаточная дискретизация видеосигнала.

Можно рекомендовать провести дополнительные исследования, касающиеся проверки четкости мониторов с ЭЛТ.

2.1.4 Контрастность монитора

Существенное влияние на контрастность может оказывать освещенность окружающей среды.

В профессиональных мониторах с ЭЛТ 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 устройствах с ЭЛТ используются технологии для получения большей контрастности в условиях сильной освещенности.

Для расчета контрастности данной ЭЛТ необходимо знать коэффициент отражения экрана K такой ЭЛТ. В лучшем случае коэффициент отражения экрана приблизительно составляет $K = 6\%$.

При рассеянной окружающей освещенности $I = 200$ люкс и $K = 6\%$ (3,82 кд/м²) яркость отражения неактивных областей экрана рассчитывается по следующей формуле:

$$L_{reflected} = \frac{I}{\pi} K .$$

При заданных значениях яркость отражения (кд/м²) примерно составляет 2% от падающей освещенности (люкс).

Считается, что ЭЛТ не дает зеркальных отражений от переднего стекла, точное влияние которых на контрастность трудно оценить количественно ввиду большой зависимости от условий освещения.

В пп. 2.1.1 и 2.1.2 коэффициент контрастности CR выражается как:

$$CR = L_{min} / L_{max} ,$$

где:

L_{min} : яркость неактивных областей при окружающей освещенности (при заданных значениях $L_{min} = L_{inactive\ areas} + L_{reflected} = 3,82$ кд/м²);

L_{max} : яркость белых областей при окружающей освещенности (кд/м²) (при заданных значениях $L_{max} = L_{white} + L_{reflected} = 200 + 3,82$ кд/м²).

При таких величинах получаем значение $CR = 0,018$, весьма близкое к значению 0,02, указанному в пп. 2.1.1.1 и 2.1.2.1, а).

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.

Изображения и последовательности сигналов, сохраняемые в цифровой форме, могут быть воспроизведены наибольшее число раз, и, следовательно, они являются предпочтительными. Лаборатории могут обменяться такими сигналами для того, чтобы сделать сравнение систем более значимым. Возможными форматами могут служить видеолента или лента для компьютера.

Короче говоря, 35-мм сканеры слайдов обеспечивают предпочтительный исходный сигнал для неподвижных изображений. 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.

Зачастую требуется учесть, каким образом параметры испытываемой системы могли бы быть изменены в ходе обработки, выполненной на более ранних стадиях существования сигнала. Следовательно, желательно, чтобы тестирование 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 цифровом 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 тестового материала для использования, при необходимости, в будущем, и вместе с ними должно храниться подробное описание истории записанного сигнала.

2.3 Выбор испытательных материалов

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

ТАБЛИЦА 1

Выбор испытательных материалов*

Задачи оценки	Используемый материал
Общее качество со средним материалом	Обычный, "критичный, но не чрезмерно"
Пропускная способность, критичные применения (например, подача программ, пост-обработка и т. д.)	Набор, включающий очень критичный материал для тестируемых 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 тестовые материалы могут быть частью содержания телевизионной программы. Более полное руководство по выбору тестовых материалов содержится в Дополнениях 1 и 2 к Приложению 1.

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 тестовых последовательностей для всего времени демонстрации программ. Без знания формы этого распределения, что обычно имеет место, следует очень осторожно производить отбор тестового материала и толкование результатов.

Как правило, очень важно включать критичный материал, т. к. его можно учесть при толковании результатов, однако невозможно осуществлять экстраполяцию исходя из некритичных материалов. В тех случаях, когда содержание сцены или последовательности изображений влияет на результаты, для испытуемой 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.

Многие организации разработали тестовые неподвижные изображения и последовательности. Существует надежда на организацию этой работы в будущем в рамках МСЭ. В рекомендациях, предназначенных для оценки применений, предлагается специальный материал по изображениям.

Дополнительные соображения по отбору испытательных материалов приведены в Дополнениях 1 и 2 к Приложению 1.

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

Необходимо, чтобы наблюдателей было не менее 15. Они не должны быть экспертами, что означает, что они не связаны непосредственно с оценкой качества телевизионного изображения, как частью их обычной работы, и они не являются опытными оценщиками (см. Примечание 1). До сеанса следует оценить (или скорректировать) остроту зрения наблюдателей при помощи таблицы Снеллена и Ландольта и проверить цветовое зрение, используя специально отобранные таблицы (например, Ишихара). Необходимое количество наблюдателей зависит от критичности и надежности принятой процедуры испытания и от ожидаемого воздействия оцениваемого эффекта.

ПРИМЕЧАНИЕ 1. – По предварительным заключениям с помощью наблюдателей, не являющихся экспертами, могут быть получены более ценные результаты при оценке технологий передачи и устройств отображения более высокого качества.

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 существовании различных уровней навыков разных групп оценщиков, не являющихся экспертами. Необходимо провести дополнительное исследование для оценки обоснованности этой гипотезы, и, если она окажется верной, количественно оценить изменения, обусловленные данным фактором. Однако на промежуточном этапе экспериментаторы должны как можно более подробно указывать характеристики своих групп оценки с целью содействия изучению этого фактора. Рекомендуемые для предоставления данные могут включать: профессиональную категорию (например, работник радиовещательной организации, студент университета, служащий), пол и возрастную группу.

2.6 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.

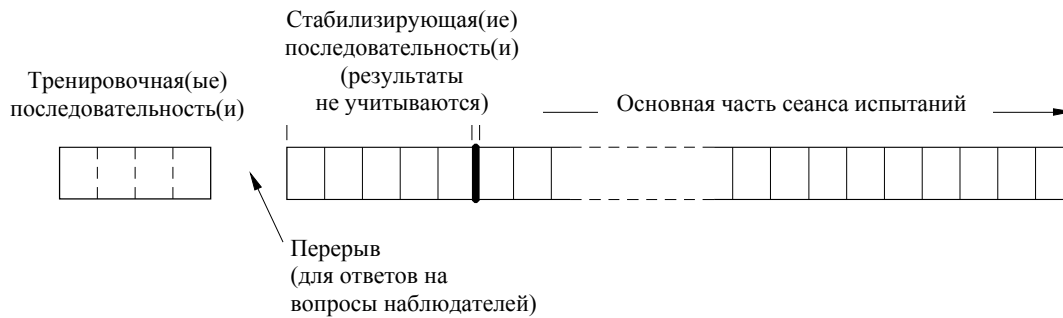
2.7 Сеанс испытаний

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 клеток); 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

Структура представления сеанса испытаний



0500-01

2.8 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 2 к данной 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 мониторов (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. Обзор типичных задач по оценке и методов, используемых для их решения, приведен в таблице 2.

ТАБЛИЦА 2
Выбор методов испытаний

Задача оценки	Используемый метод	Описание
Измерение качества систем по отношению к эталонному	Метод с двумя источниками воздействия и непрерывной шкалой качества и с использованием шкалы искажений (DSCQS) ⁽¹⁾	Рек. МСЭ-R ВТ.500, п. 5
Измерение устойчивости систем (т. е. характеристики отказов)	Метод оценки искажений с двумя источниками воздействия (DSIS) ⁽¹⁾	Рек. МСЭ-R ВТ.500, п. 4
Количественная оценка качества систем (когда эталон отсутствует)	Метод относительного масштабирования ⁽²⁾ или категорического масштабирования (исследуется)	Отчет МСЭ-R ВТ.1082
Сравнение качества альтернативных систем (когда эталон отсутствует)	Метод прямого сравнения, метод относительного масштабирования ⁽²⁾ или категорического масштабирования (исследуется)	Отчет МСЭ-R ВТ.1082
Определение факторов, по которым системы воспринимаются как разные, и измерение их влияния на восприятие	Исследуемые методы	Отчет МСЭ-R ВТ.1082
Установление точки, когда искажение становится видимым	Оценка порога при помощи метода принудительного выбора или метода подстройки (исследуется)	Отчет МСЭ-R ВТ.1082
Определение того, воспринимаются ли системы как разные	Метод принудительного выбора (исследуется)	Отчет МСЭ-R ВТ.1082
Измерение качества кодирования стереоскопического изображения	Метод с двумя источниками воздействия и непрерывной шкалой качества (DSCQS) ⁽³⁾	Рек. МСЭ-R ВТ.500, п. 5
Количественная оценка верности между двумя видеопоследовательностями с ухудшенным качеством	Метод с двумя источниками одновременного воздействия для непрерывной оценки (SDSCE)	Рек. МСЭ-R ВТ.500, п. 6.4
Сравнение различных средств обеспечения устойчивости к ошибкам	Метод с двумя источниками одновременного воздействия для непрерывной оценки (SDSCE)	Рек. МСЭ-R ВТ.500, п. 6.4

(1) Был проведен ряд исследований контекстуальных эффектов в отношении методов DSCQS и DSIS. Было обнаружено, что результаты, полученные с помощью метода DSIS, в некоторой степени подвержены влиянию контекстуальных эффектов. Более подробная информация приведена в Дополнении 3 к Приложению 1.

(2) В ряде исследований утверждается, что этот метод наиболее стабилен в том случае, когда доступен полный диапазон качества.

(3) Ввиду возможности большой усталости при оценивании стереоскопических изображений общая длительность сеанса испытаний должна быть снижена менее чем до 30 минут.

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

4.1 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 п. 5.

Метод с двумя источниками воздействия (ЕСР) – циклический, в нем эксперту по оценке 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, эксперту 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 тестового 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.

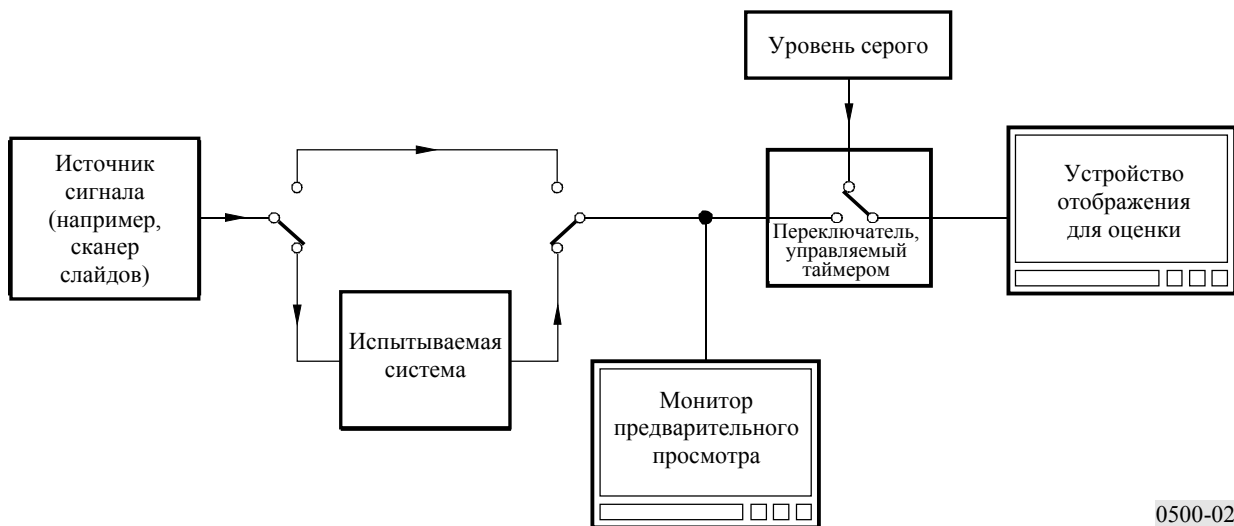
4.2 Общий порядок

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

Общий набор оборудования для испытательной установки должен быть таким, как показано на рис. 2.

РИСУНОК 2

Общая компоновка испытательной системы для метода DSIS



0500-02

Эксперты 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. Экспертам 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.

4.3 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 рис. 3а.

Вариант II: Эталонное изображение или последовательность и оцениваемое изображение или последовательность представляются дважды, как это показано на рис. 3б.

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

4.4 Шкалы оценок

Следует использовать пятибалльную шкалу оценки искажений:

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

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

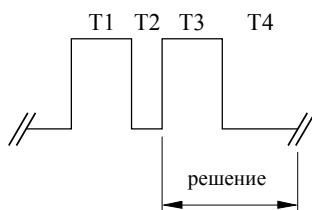
4.5 Представление оценок

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 выразить свое решение терминами формулировок, используемых для установления субъективной шкалы.

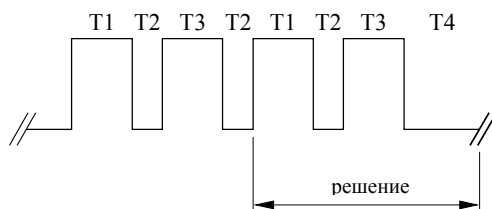
Наблюдателей следует просить рассматривать изображение в течение полных периодов T1 и T3. Принятие решения должно быть разрешено только в течение T4.

РИСУНОК 3

Структура представления материала для испытаний



а) Вариант I



б) Вариант II

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 экспертов давать правильную оценку изображениям или последовательностям.

4.6 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 Дополнении 2 к Приложению 1.

5 Метод двух источников воздействия с непрерывной шкалой качества (DSCQS)

5.1 Общее описание

Типичный метод может быть применен для оценки либо новой системы, либо влияния 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 эксперта 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, эксперту 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.

5.2 Общий порядок

Условия просмотра, исходные сигналы, материал для испытаний, наблюдатели и ознакомление с оценкой определяются или выбираются в соответствии с п. 2. Сеанс испытаний описан в п. 4.6.

Общий набор оборудования для испытательной системы должен быть таким, как показано на рис. 4.

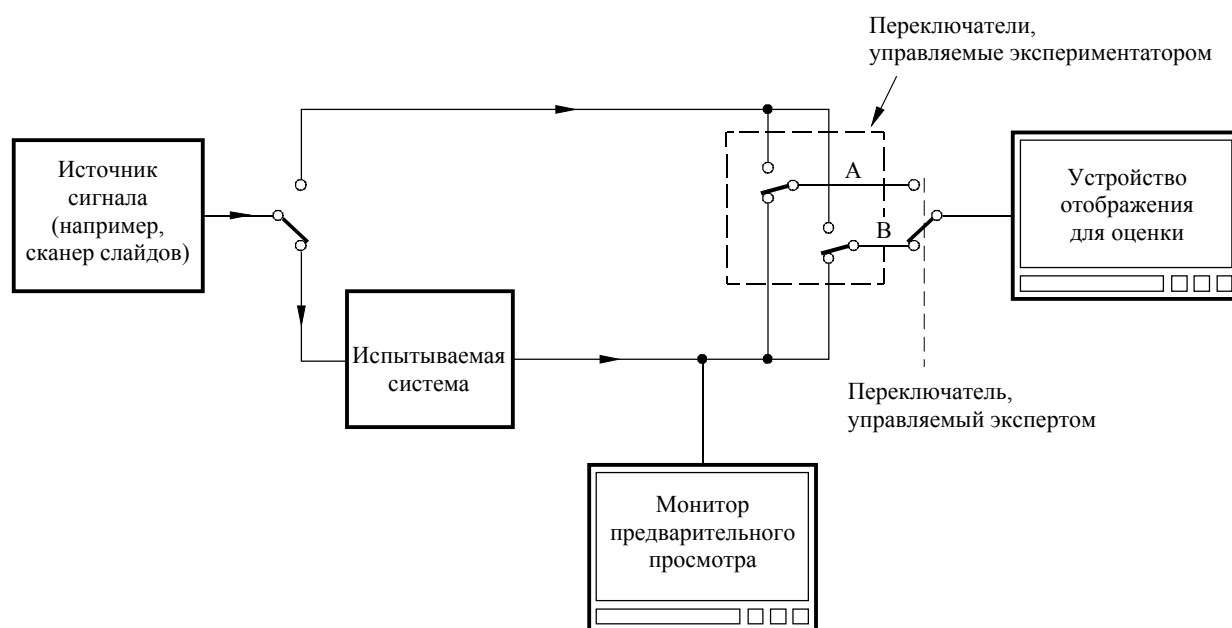
5.3 Представление материала для испытаний

Сеанс испытаний состоит из ряда представлений. В случае варианта I, при котором участвует один наблюдатель, во время каждого представления эксперт может просматривать сигналы А и В до тех пор, пока он не оценит качество каждого сигнала. Эксперт обычно может сделать это 2 или 3 раза за период 10 с. В случае варианта II, в котором одновременно участвуют несколько экспертов, перед регистрацией результатов пара изображений демонстрируется один или несколько раз в течение равных промежутков времени для того, чтобы эксперт мог мысленно оценить их качества, затем эта пара демонстрируется еще один или несколько раз, при этом результаты регистрируются. Число повторений зависит от длительности тестовых последовательностей. Для неподвижных изображений могут быть целесообразными 3–4 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 рис. 5.

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-ти 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 4

Общая компоновка испытательной системы для метода DSCQS



Существует два варианта данного метода (I) и (II), описанные в общих чертах 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 подаются 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, отмечается экспериментатором, но не объявляется.

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 подаются так же, как было описано выше для варианта (I). Стабильность результатов при этом варианте с ограниченным диапазоном изменения качества все еще изучается.

0500-04

5.4 Шкала оценок

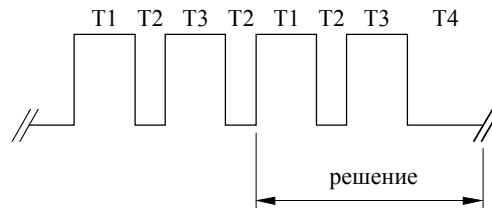
Данный метод требует оценки обоих вариантов каждого испытательного изображения. В каждой паре одно испытательное изображение является неискаженным, тогда как другое представление может либо содержать, либо не содержать искажений. Неискаженное изображение включается для того, чтобы служить эталоном, но 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 двоякого 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 рис. 6 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 5

Структура представления материала для испытания



Фазы представления:

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

0500-05

РИСУНОК 6

Часть формы для оценки качества с использованием непрерывных шкал*

	27		28		29		30		31	
	A	B	A	B	A	B	A	B	A	B
Отлично										
Хорошо										
Удовлетворительно										
Плохо										
Неприемлемо										

* При планировании порядка представления тестового 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.

0500-06

5.5 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.

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 тестовой 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. Дополнение 1 к Приложению 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.

5.6 Толкование 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.

6 Альтернативные методы оценки

При определенных обстоятельствах должны быть использованы методы с одним источником воздействия и методы сравнения воздействий.

6.1 Методы с одним источником воздействия (SS)

В методах с одним источником воздействия демонстрируется одно изображение или последовательность изображений, и эксперт 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.

6.1.1 Общий порядок

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

6.1.2 Выбор материала для испытаний

Содержание испытательных изображений для лабораторных испытаний должно выбираться, как описано в п. 2.3.

После выбора содержания испытательные изображения подготавливаются 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 (т. е. неаддитивные влияния).

6.1.3 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 п. 2.7); 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 зрителя 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 мониторе 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.

6.1.4 Типы методов с одним источником воздействия (SS)

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

6.1.4.1 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 3. 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 3

МСЭ-R шкалы 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.

6.1.4.2 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.

6.1.4.3 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 – эксперт присваивает каждому изображению или последовательности изображений точку на линии, проведенной между двумя семантическими метками (т. е. 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 эксперт 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").

6.1.4.4 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.

6.2 Методы сравнения воздействий

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

6.2.1 Общий порядок

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

6.2.2 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.

6.2.3 Сеанс испытаний

В эксперименте по оценке будет использоваться либо один, либо два монитора с хорошим качеством, и он будет проходить, как правило, аналогично случаю с одним источником воздействия. Если используется один монитор, 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 монитора, 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 экспертами или использование выборки из всех возможных пар.

6.2.4 Типы методов со сравнением воздействий

В телевизионных оценках используется три типа методов со сравнением воздействий.

6.2.4.1 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 4.

ТАБЛИЦА 4

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.).

6.2.4.2 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 эксперт присваивает каждому соотношению цифру на линии, проведенной между двумя метками (например, ТО ЖЕ – ОТЛИЧНЫЙ 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 эксперт 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.

6.2.4.3 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.

6.3 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 следующая новая методика SSCQE.

6.3.1 Непрерывная оценка общего качества

6.3.1.1 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.

6.3.1.2 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.

6.3.1.3 Параметры просмотра

Условия просмотра должны быть условиями, указанными в настоящее время в Рекомендациях МСЭ-R ВТ.500, ВТ.1128, ВТ.1129 и ВТ.710.

6.3.1.4 Шкалы оценок

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

6.3.1.5 Наблюдатели

Не менее пятнадцати участников, не являющиеся экспертами, должны быть использованы при условиях, в настоящее время рекомендованных в п. 2.5.

6.3.1.6 Инструкции для наблюдателей

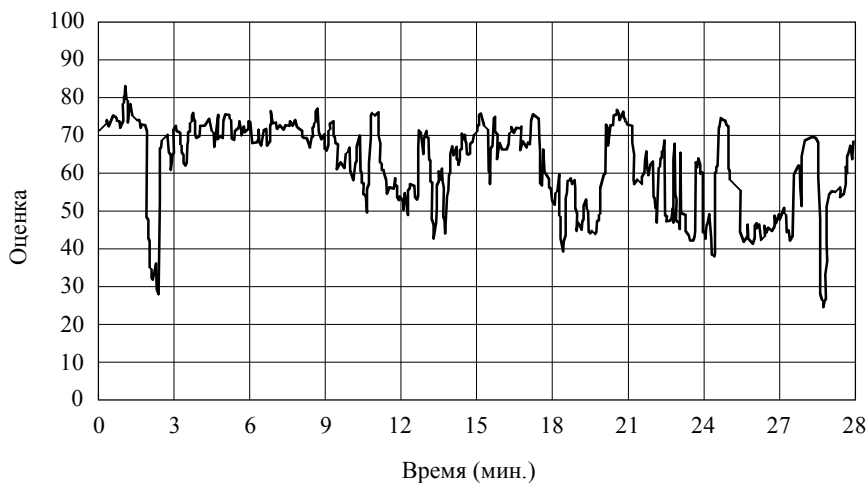
В случае оценки качества обслуживания (при звуковом сопровождении) наблюдатели должны быть проинструктированы о том, что они должны учитывать общее качество, а не только качество видеоизображения.

6.3.1.7 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 рис. 7).

РИСУНОК 7

Условие испытания. Кодек X/Фрагмент программы: Z



0500-07

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 рис. 8).

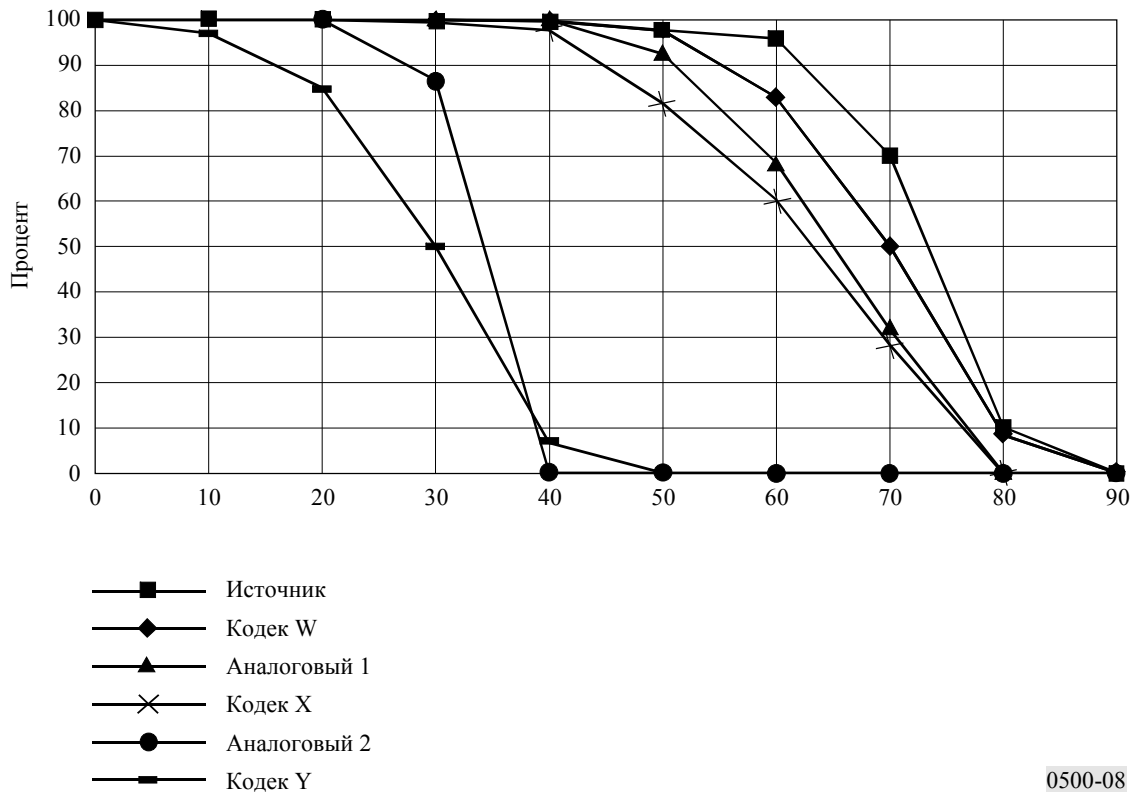
6.3.2 Калибровка результатов непрерывной оценки качества и получение общей оценки качества

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 с, 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 8

Средние оценки последовательностей фрагмента программы Z,
по которым проводится голосование



0500-08

6.4 Метод непрерывной оценки с двумя источниками одновременного воздействия

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 методов 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.

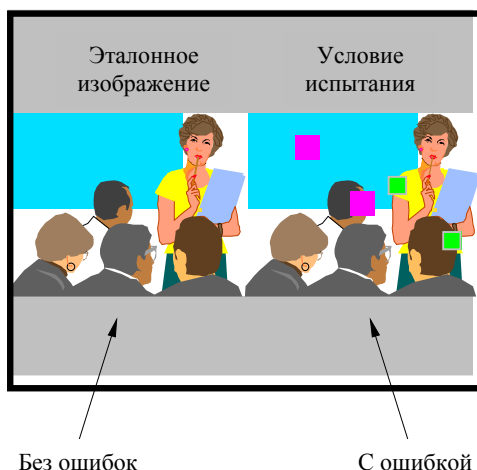
В результате был разработан и испытан следующий новый метод SDSCE.

6.4.1 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 мониторе, в ином случае должны использоваться два рядом стоящих монитора (см. рис. 9).

РИСУНОК 9

Пример формата устройства отображения



0500-09

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.

6.4.2 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" (т. е. разница не была замечена); если же вместо этого участники заявляют о том, что видят некоторые различия, экспериментатор должен повторить объяснение и моделирование испытания.

6.4.3 Особенности протокола испытаний

Следующие определения применяются к описанию протокола испытаний:

- *Фрагмент видеоизображения (ФВ)*: ФВ соответствует одной последовательности видеоизображений.
- *Условие испытания (УИ)*: УИ может быть конкретный процесс, связанный с видеоизображением, 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)*: десятисекундный сегмент вынесения решений; все СПР получают путем использования групп 20 последовательных решений (эквивалентных 10 секундам) без какого-либо перекрытия.

6.4.4 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.

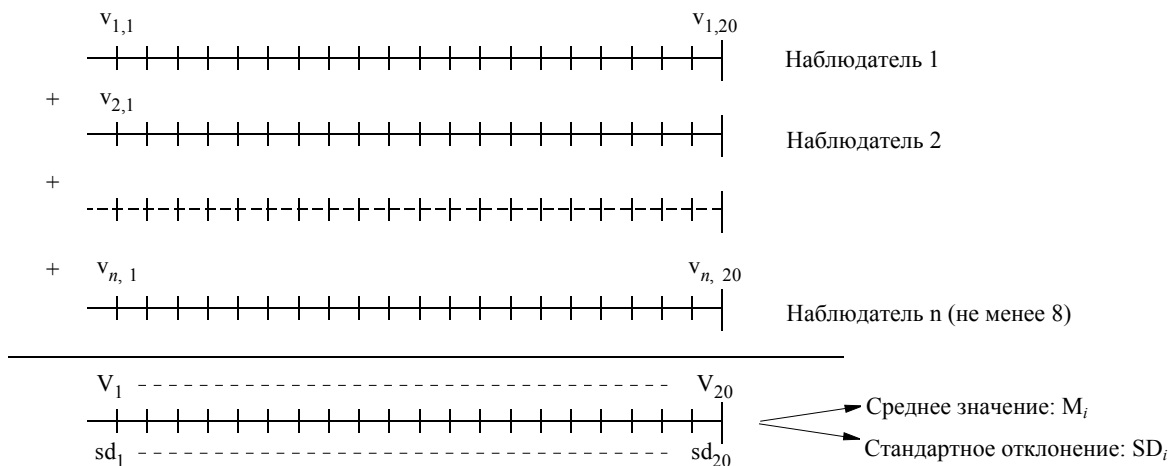
В каждом случае необходим многошаговый анализ:

- Средние значения и стандартные отклонения рассчитываются по каждому решению путем накопления наблюдателей.
- Средние значения и стандартное отклонение рассчитываются для каждого СПР, как показано на рис. 10. Результаты этого этапа могут быть представлены на временной диаграмме, как показано на рис. 11.
- Анализируется статистическое распределение средних значений, рассчитанных на предыдущем этапе (т. е. 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 рис. 12. 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.

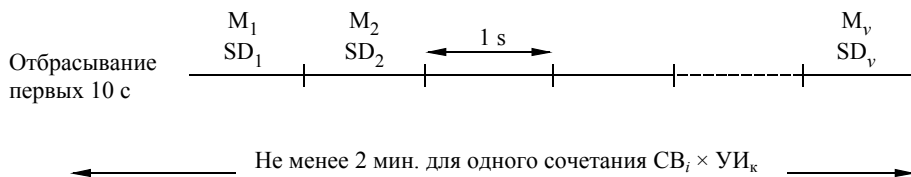
РИСУНОК 10

Обработка данных

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



- б) Расчет M и SD на последовательность для голосования продолжительностью 1 с для каждого сочетания $VS \times TS$



0500-10

6.4.5 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 п. 2.3.2 Приложения 2.

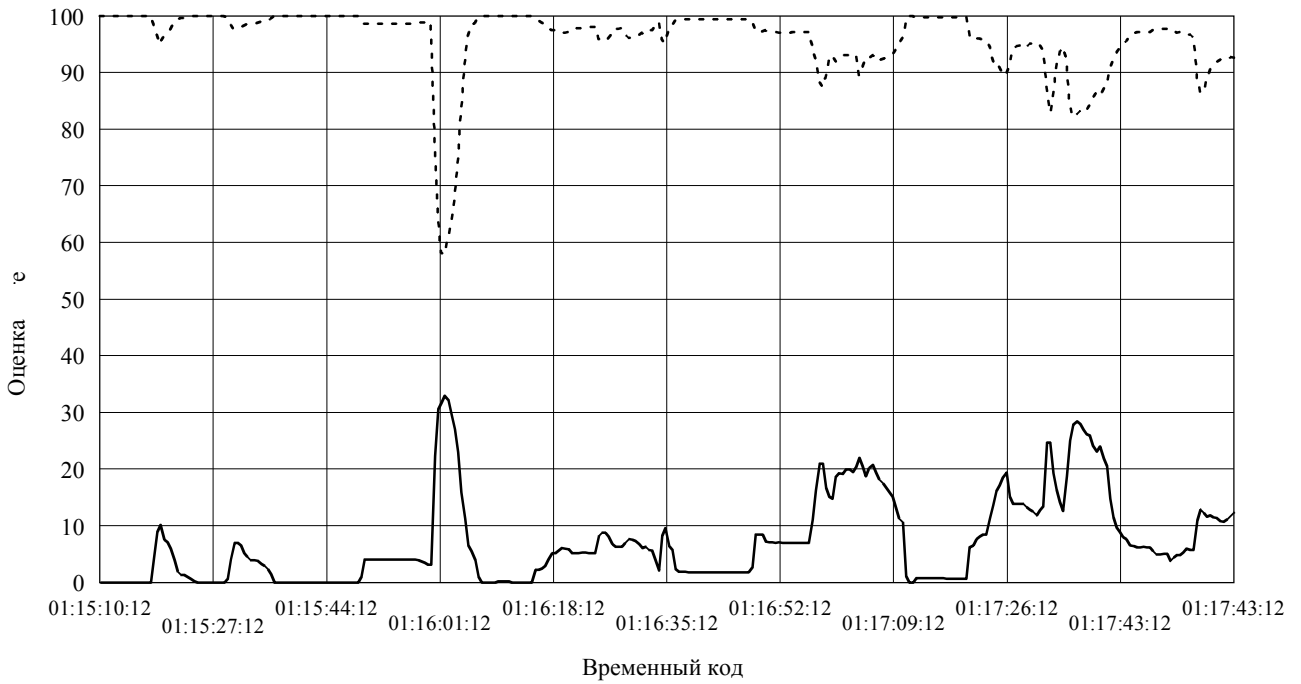
В случае процедуры SDSCE надежность решений зависит от следующих двух параметров:

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.

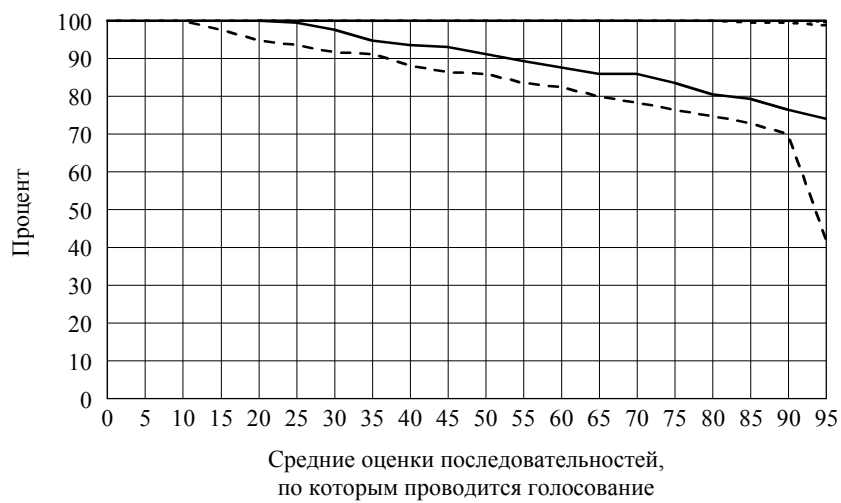
РИСУНОК 11
Групповая временная диаграмма



----- Среднее значение
 ————— Стандартное отклонение

0500-11

РИСУНОК 12
Общие характеристики раздражающего воздействия, рассчитанные на основе статистических распределений и включающие доверительный интервал



----- Критическое
 ————— Нет ошибок
 - - - - - Типичное

0500-12

6.5 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 обстоятельствам, остается на усмотрение исследователя.

Ограничения, присущие различным методам, наводят на мысль, что может быть неразумно слишком полагаться на один метод. Следовательно, может быть целесообразно рассмотреть более "полные" подходы, такие как использование нескольких методов или использование многомерного подхода.

Дополнение 1 к Приложению 1

Характеристики искажений содержания изображения

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.

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 рис. 13.

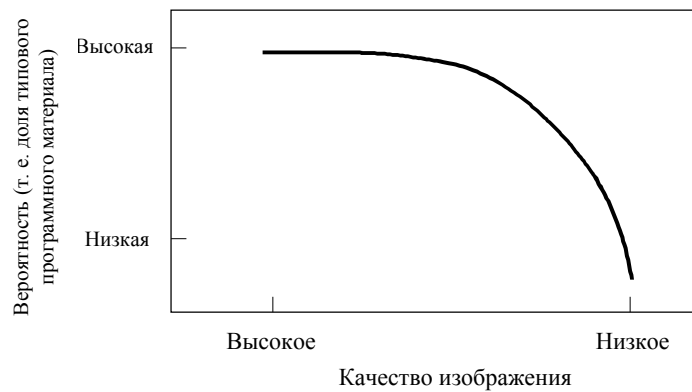
Характеристика искажения изображения может быть получена за четыре шага:

- *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 рис. 14.

- *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 рис. 15.
- *Шаг 4:* заключается в объединении информации, полученной на шагах 2 и 3 для того, чтобы получить характеристику искажений содержания изображения в форме, представленной на рис. 13.

РИСУНОК 13

Графическое представление возможной характеристики искажения изображений



0500-13

РИСУНОК 14

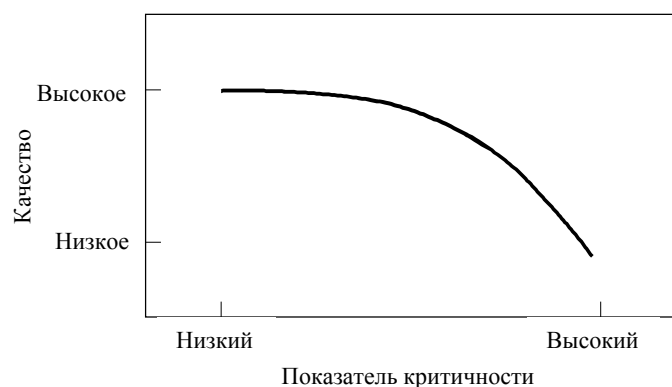
Вероятность появления материала с различными уровнями критичности



0500-14

РИСУНОК 15

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



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 (т. е. для того, чтобы предвидеть масштаб и серьезность искажений во время работы);
- для оценки ее 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 Рекомендации МСЭ-R ВТ.1129.

Дополнение 2 к Приложению 1

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

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.

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 Дополнение 1 к Приложению 1). Этот результат будет объединен с результатами анализа условий передачи для подготовки к завершающему этапу процесса.

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.

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. п. 5 Приложения 1). 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.

Дополнение 3 к Приложению 1

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 2

Анализ и представление результатов

1 Введение

В ходе субъективного эксперимента по оценке качества телевизионной системы осуществляется сбор большого объема данных. Эти данные в форме бланков экспертных оценок 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 настоящей Рекомендации (п. 4, 5 и 6 в Приложении 1), и к другим альтернативным методам, в которых применяются цифровые шкалы. В первом и втором случаях используются непрерывные шкалы оценок, а результаты (различия оценок эталонного изображения и испытываемого реального изображения) приводятся к целочисленным значениям от 0 до 100.

2 Общие методы анализа

В результате проведения испытаний в соответствии принципами методов, которые описаны в Приложении 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 .

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.

2.2 Расчет доверительного интервала

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-процентный доверительный интервал, который задается выражением:

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

где

$$\delta_{jkr} = 1,96 \frac{S_{jkr}}{\sqrt{N}}. \quad (2)$$

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

$$S_{jkr} = \sqrt{\frac{\sum_{i=1}^N (\bar{u}_{jkr} - u_{ijkr})^2}{(N-1)}}. \quad (3)$$

С вероятностью 95% абсолютное значение разницы между экспериментальной средней оценкой и "верной" средней оценкой (при очень большом числе наблюдателей) меньше 95-процентного 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 экспертами, участвующими в оценке.

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.

2.3 Отбор наблюдателей

2.3.1 Отбор 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, т. е. 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_{ijk} 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 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 1. – 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 ECP (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:

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \quad \text{с} \quad m_x = \frac{\sum_{i=1}^N (u_{ijk} - \bar{u}_{jkr})^x}{N}. \quad (4)$$

Для каждого наблюдателя i находят P_i и Q_i , т. е.:

для $j, k, r = 1, 1, 1$ до J, K, R

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

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

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

или же:

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

если $u_{ijkkr} \leq \bar{u}_{jkr} - \sqrt{20} S_{jkr}$, 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).

2.3.2 Отбор в случае 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}_{jklr} , стандартное отклонение S_{jklr} и коэффициент β_{2jklr} , с использованием оценок u_{ijkkr} каждого наблюдателя. Коэффициент β_{2jklr} задается формулой:

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

Для каждого наблюдателя i находят P_i и Q_i , т. е.:

для $j, k, l, r = 1, 1, 1, 1$ 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 рис. 17 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 к настоящей Рекомендации. 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:

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

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

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

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:

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

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:

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

Для каждого наблюдателя i находят P^*_i и Q^*_i , т. е.:

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.

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 п. 2.1), то может быть полезным следующий процесс, заключающийся в нахождении простой непрерывной зависимости между \bar{u} и параметром искажений.

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

$$p = (\bar{u} - u_{min}) / (u_{max} - u_{min}), \quad (5)$$

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:

$$p = 1/[1 + \exp(D - D_M) G], \quad (6)$$

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:

$$I = (1/p - 1). \quad (7)$$

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

$$I = \exp(D - D_M) G. \quad (8)$$

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

$$\log_e I = (D - D_M) G. \quad (9)$$

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:

$$S = \frac{D_M - D}{\log_e I} = \frac{1}{G}, \quad (10)$$

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.

3.2 Аппроксимация несимметричной функцией

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 (8) следует заменить формулой:

$$I = (d/d_M)^{1/G}, \quad (11)$$

и поэтому (6) принимает вид:

$$p = 1/[1 + (d/d_M)^{1/G}]. \quad (12)$$

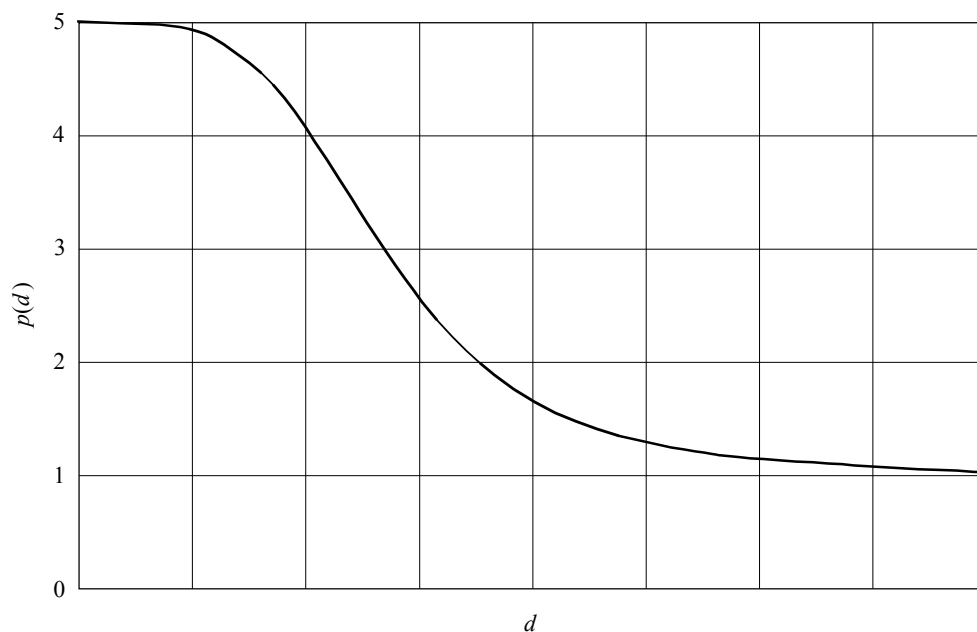
Данная функция аппроксимирует логическую функцию несимметричным способом.

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 рис. 16 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 16

Несимметричная аппроксимация



0500-16

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 5).

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 5

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	Да
Преобразование линейной шкалы	Да	Может быть существенная ошибка	Да
Преобразование нелинейной шкалы ⁽¹⁾	Да	Да	Да
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.

3.4 Включение аспекта надежности в графики

Из средних оценок для каждого испытанного искажения и соответствующих 95-процентных 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. рис. 17). Следовательно, мы можем получить оценку 95-процентной непрерывной доверительной области.

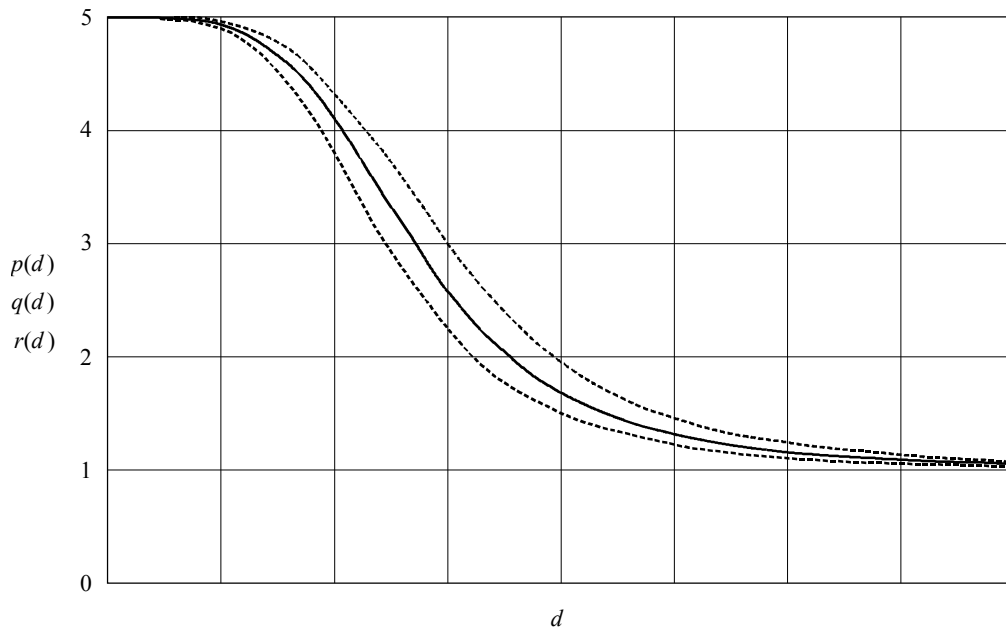
Таким образом, для оценки 4,5 (порог заметности для данного метода) мы можем непосредственно из графика найти 95-процентный доверительный интервал, который может быть использован с целью определения диапазона допусков.

Промежуток между кривыми для максимальных и минимальных оценок – это не 95-процентный 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 17

Случай несимметричной характеристики искажений



$p(d)$: ряд средних оценок
 $q(d)$: ряд минимальных оценок
 $r(d)$: ряд максимальных оценок
 d : объективная количественная оценка искажений

0500-17

4 Выводы

Описана процедура оценки доверительных интервалов, т.е. 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.

Приложение 3

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 толкование 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 – критического, 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 Рекомендации МСЭ-R ВТ.500.

Он состоит из текстовых файлов, имеющих структуру, которая представлена в таблицах 6 и 7. 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 6

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 = целочисленное значение↵ Максимальное значение по шкале = целочисленное значение↵Scale maximum = integer↵ Размер монитора = целочисленное значение↵ Марка и модель монитора = Последовательность знаков↵ [РЕЗУЛЬТАТЫ]↵ Количество результатов = $1 \leq \text{целочисленное значение} \leq y$ ↵ Результат(j). Название файла = Последовательность знаков.DAT↵ Результат(j). Название = Последовательность знаков↵ Результат(j). Лаборатория = Последовательность знаков↵ Результат(j). Число наблюдателей = $1 \leq \text{целочисленное значение} \leq N$ ↵ Результат(j). Обучение = "Да" или "Нет" ↵ [Результат(j). Сеанс(i). Наблюдатели] ↵ O(k). Имя = Последовательность знаков↵ O(k). Фамилия = Последовательность знаков↵ O(k). = "Ж" или "М" ↵ O(k). Возраст = целочисленное значение↵ O(k). Профессия = целочисленное значение↵ O(k). Расстояние = целочисленное значение↵	[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 [Идентификатор раздела] Идентификация наблюдателя Дополнительно Дополнительно Основные социально-экономические группы (например, рабочий, студент) Расстояние просмотра при высотах устройства отображения (например, 3 H, 4 H, 6 H)

⁽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 7

Формат текстового файла первоначальных данных Результаты.DAT

Формат и синтаксис файла название файла.DAT	Замечания
целое целое целое.....↵ целое целое целое.....↵ 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(j). Название(я) файла(ов) ⁽¹⁾ .

⁽¹⁾ См. сноску⁽¹⁾ к таблице 6.