

Международный союз электросвязи

МСЭ-Т

СЕКТОР СТАНДАРТИЗАЦИИ
ЭЛЕКТРОСВЯЗИ МСЭ

P.910

(04/2008)

СЕРИЯ Р: КАЧЕСТВО ТЕЛЕФОННОЙ ПЕРЕДАЧИ,
ТЕЛЕФОННЫЕ УСТАНОВКИ, СЕТИ МЕСТНЫХ
ЛИНИЙ

Аудиовизуальное качество в мультимедийных услугах

**Методы субъективной оценки качества
видеоизображения для мультимедийных
приложений**

Рекомендация МСЭ-Т Р.910



РЕКОМЕНДАЦИИ МСЭ-Т СЕРИИ Р

КАЧЕСТВО ТЕЛЕФОННОЙ ПЕРЕДАЧИ, ТЕЛЕФОННЫЕ УСТАНОВКИ, СЕТИ МЕСТНЫХ ЛИНИЙ

Словарь и воздействие параметров передачи на мнение клиента о качестве передачи	Серия	P.10
Характеристики голосовых терминалов	Серия	P.30 P.300
Эталонные системы	Серия	P.40
Аппаратура для объективных измерений	Серия	P.50 P.500
Объективные электроакустические измерения	Серия	P.60
Измерения, относящиеся к громкости речи	Серия	P.70
Методы объективной и субъективной оценки качества речи	Серия	P.80
Методы объективной и субъективной оценки качества речи и видео	Серия	P.800
Аудиовизуальное качество в мультимедийных услугах	Серия	P.900
Аспекты качества передачи и QoS в конечных точках IP-сетей	Серия	P.1000
Связь в транспортных средствах	Серия	P.1100
Модели и средства для оценки качества потокового мультимедиа	Серия	P.1200
Оценка телесобраний	Серия	P.1300
Руководящие указания по статистическому анализу, оценке и отчетности для измерений качества	Серия	P.1400
Методы объективной и субъективной оценки качества услуг, отличных от услуг передачи речи и видео	Серия	P.1500

Для получения более подробной информации просьба обращаться к перечню Рекомендаций МСЭ-Т.

Рекомендация МСЭ-Т Р.910

Методы субъективной оценки качества видеозображения для мультимедийных приложений

Резюме

В настоящей Рекомендации МСЭ-Т Р.910 описаны неинтерактивные субъективные методы оценки общего качества при односторонней передаче видеозображений для мультимедийных приложений, таких как видеоконференц-связь, хранение и извлечение видеозаписей, телемедицина и т. п. Эти методы можно использовать для разнообразных целей, в том числе для выбора алгоритмов, классификации рабочих характеристик системы и оценки уровня качества во время видеосвязи. В данной Рекомендации также описаны характеристики исходных последовательностей, которые предполагается использовать, такие как длительность, тип контента, количество последовательностей и т. д.

Хронологическая справка

Издание	Рекомендация	Утверждение	Исследовательская комиссия	Уникальный идентификатор [*]
1.0	МСЭ-Т Р.910	30.08.1996 г.	12-я	11.1002/1000/3641
2.0	МСЭ-Т Р.910	30.09.1999 г.	12-я	11.1002/1000/4751
3.0	МСЭ-Т Р.910	06.04.2008 г.	9-я	11.1002/1000/9317

^{*} Для получения доступа к Рекомендации наберите в адресном поле вашего веб-браузера URL: <http://handle.itu.int/>, после которого укажите уникальный идентификатор Рекомендации. Например, <http://handle.itu.int/11.1002/1000/11830-en>.

ПРЕДИСЛОВИЕ

Международный союз электросвязи (МСЭ) является специализированным учреждением Организации Объединенных Наций в области электросвязи и информационно-коммуникационных технологий (ИКТ). Сектор стандартизации электросвязи МСЭ (МСЭ-Т) – постоянный орган МСЭ. МСЭ-Т отвечает за изучение технических, эксплуатационных и тарифных вопросов и за выпуск Рекомендаций по ним в целях стандартизации электросвязи на всемирной основе.

На Всемирной ассамблее по стандартизации электросвязи (ВАСЭ), которая проводится каждые четыре года, определяются темы для изучения исследовательскими комиссиями МСЭ-Т, которые в свою очередь вырабатывают Рекомендации по этим темам.

Утверждение Рекомендаций МСЭ-Т осуществляется в соответствии с процедурой, изложенной в Резолюции 1 ВАСЭ.

В некоторых областях информационных технологий, которые входят в компетенцию МСЭ-Т, необходимые стандарты разрабатываются на основе сотрудничества с ИСО и МЭК.

ПРИМЕЧАНИЕ

В настоящей Рекомендации термин "администрация" используется для краткости и обозначает как администрацию электросвязи, так и признанную эксплуатационную организацию.

Соблюдение положений данной Рекомендации осуществляется на добровольной основе. Однако данная Рекомендация может содержать некоторые обязательные положения (например, для обеспечения функциональной совместимости или возможности применения), и в таком случае соблюдение Рекомендации достигается при выполнении всех указанных положений. Для выражения требований используются слова "следует", "обязан" (shall) или некоторые другие обязывающие выражения, такие как "должен" (must), а также их отрицательные формы. Употребление таких слов не означает, что от какой-либо стороны требуется соблюдение положений данной Рекомендации.

ПРАВА ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

МСЭ обращает внимание на вероятность того, что практическое применение или выполнение настоящей Рекомендации может включать использование заявленного права интеллектуальной собственности. МСЭ не занимает какую бы то ни было позицию относительно подтверждения, действительности или применимости заявленных прав интеллектуальной собственности независимо от того, доказываются ли такие права членами МСЭ или другими сторонами, не относящимися к процессу разработки Рекомендации.

На момент утверждения настоящей Рекомендации МСЭ не получил извещение об интеллектуальной собственности, защищенной патентами, которые могут потребоваться для выполнения настоящей Рекомендации. Однако те, кто будет применять Рекомендацию, должны иметь в виду, что вышесказанное может не отражать самую последнюю информацию, и поэтому им настоятельно рекомендуется обращаться к патентной базе данных БСЭ по адресу <http://www.itu.int/ITU-T/ipr/>.

© ITU 2009

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.

СОДЕРЖАНИЕ

	Стр.
1 Сфера применения	1
2 Справочные документы	1
3 Термины и определения	1
4 Сокращения и акронимы	3
5 Исходный сигнал.....	4
5.1 Условия записи	4
5.2 Система записи	4
5.3 Характеристики сцены	5
6 Методы испытаний и планирование эксперимента	6
6.1 Показатель абсолютной категории (ACR).....	6
6.2 Показатель абсолютной категории со скрытым эталоном (ACR-HR).....	7
6.3 Оценка ухудшения категории качества (DCR)	8
6.4 Метод парного сравнения (PC).....	9
6.5 Сравнение методов	10
6.6 Эталонные условия.....	11
6.7 Планирование эксперимента	11
7 Процедуры оценки	11
7.1 Условия просмотра.....	12
7.2 Система обработки и воспроизведения	12
7.3 Зрители.....	13
7.4 Инструктаж зрителей и учебный сеанс	13
8 Статистический анализ и представление результатов.....	13
Приложение А. Детали, связанные с характеристикой тестовых последовательностей.....	15
А.1 Фильтр Собеля	15
А.2 Как использовать SI и TI для выбора тестовой последовательности	16
А.3 Примеры	16
Приложение В. Дополнительные оценочные шкалы.....	18
В.1 Рейтинговые шкалы.....	18
В.2 Дополнительные размерности оценки.....	19
Приложение С. Одновременное представление пар последовательностей.....	21
С.1 Введение	21
С.2 Синхронизация.....	21
С.3 Условия просмотра.....	21
С.4 Демонстрация.....	21
Приложение D. Классы видеоизображений и их атрибуты.....	22

	Стр.
Дополнение I. Тестовые последовательности	23
Дополнение II. Инструкции по просмотру тестовых видеоизображений	24
II.1 ACR и ACR-HR	24
II.2 DCR	24
II.3 PC	24
Дополнение III. Одновременное воздействие двух источников для непрерывной оценки	26
III.1 Процедура испытаний	26
III.2 Этап обучения	26
III.3 Особенности протокола испытаний	26
III.4 Обработка данных	27
III.5 Надежность участников	30
Дополнение IV. Объектно ориентированная оценка	32
Дополнение V. Дополнительная шкала оценок DCR	34
Библиография	35

Методы субъективной оценки качества видеоизображения для мультимедийных приложений

1 Сфера применения

Настоящая Рекомендация предназначена для определения неинтерактивных методов субъективной оценки качества цифровых видеоизображений, кодированных с битовыми скоростями, определенными в классах TV3, MM4, MM5 и MM6, как указано в таблице D.2, для таких приложений, как видеотелефония, видеоконференц-связь и хранение и извлечение видеозаписей. Эти методы можно использовать для разнообразных целей, в том числе для выбора алгоритмов, классификации рабочих характеристик системы и оценки уровня качества во время видеосвязи.

2 Справочные документы

В перечисленных ниже Рекомендациях МСЭ-Т и других справочных документах содержатся положения, которые посредством ссылок на них в этом тексте составляют основные положения данной Рекомендации. На момент опубликования действовали указанные редакции документов. Все Рекомендации и другие справочные документы являются предметом корректировки, и стороны пришли к договоренности основываться на этой Рекомендации и стараться изыскивать возможность для использования самых последних изданий Рекомендаций и справочных документов, перечисленных ниже. Регулярно публикуется перечень действующих Рекомендаций МСЭ-Т. Ссылка на документ в рамках этой Рекомендации не дает ему как отдельному документу статуса Рекомендации.

[ITU-T J.61]	Recommendation ITU-T J.61 (1988), <i>Transmission performance of television circuits designed for use in international connections.</i>
[ITU-T P.800]	Recommendation ITU-T P.800 (1996), <i>Methods for subjective determination of transmission quality.</i>
[ITU-T P.930]	Recommendation ITU-T P.930 (1996), <i>Principles of a reference impairment system for video.</i>
[ITU-R BT.500-9]	Recommendation ITU-R BT.500-9 (1998), <i>Methodology for the subjective assessment of the quality of television pictures.</i>
[ITU-R BT.601-4]	Recommendation ITU-R BT.601-4 (1994), <i>Encoding parameters of digital television for studios.</i>
[ITU-R BT.814-1]	Recommendation ITU-R BT.814-1 (1994), <i>Specifications and alignment procedures for setting of brightness and contrast of displays.</i>
[IEC/TR 60268-13]	IEC/TR 60268-13 (1998), <i>Sound system equipment – Part 13: Listening tests on loudspeakers</i> < http://webstore.iec.ch/webstore/webstore.nsf/artnum/022890 >.

3 Термины и определения

В настоящей Рекомендации определяются следующие термины.

3.1 Гамма (gamma) – параметр, которым описывается разрешающая способность при различных ступенях уровня яркости на устройстве визуального отображения. Между яркостью экрана и напряжением входного сигнала имеется нелинейная связь, при этом напряжение возрастает до значения гамма. Для введения поправки на эту нелинейную связь в камере обычно применяется поправочный коэффициент, представляющий собой обратную функцию от гаммы. Гамма также влияет на верность воспроизведения цвета.

3.2 Тесты для оптимизации (optimization tests) – субъективные тесты, которые, как правило, проводятся во время разработки или стандартизации нового алгоритма либо системы. Цель таких тестов состоит в оценке рабочих характеристик новых инструментов для оптимизации исследуемых алгоритмов или систем.

3.3 Тесты соответствия техническим требованиям (qualification tests) – субъективные тесты, проводимые обычно в целях сравнения рабочих характеристик стандартных систем или оборудования. Такие тесты должны проводиться при условиях испытаний, как можно более точно отражающих реальные условия использования.

3.4 Пространственная воспринимаемая информация (spatial perceptual information (SI)) – величина, которая обычно указывает на объем пространственных деталей изображения. Как правило, она выше для более сложного визуализированного трехмерного пространства. Она не предназначена служить величиной энтропии или быть связанной с информацией, определяемой в теории связи. Уравнение для SI см. в пункте 5.3.1.

3.5 Временная воспринимаемая информация (temporal perceptual information (TI)) – величина, которая обычно указывает на объем временных изменений в последовательности видеок кадров. Как правило, она выше для более высоких последовательностей движения. Она не предназначена служить величиной энтропии или быть связанной с информацией, определяемой в теории связи. Уравнение для TI см. в пункте 5.3.2.

3.6 Прозрачность (точность воспроизведения) (transparency (fidelity)) – понятие, описывающее работу кодека или системы по отношению к идеальной системе передачи без какого-либо ухудшения качества. Могут быть определены два типа прозрачности.

Первый тип описывает с использованием математических критериев, в какой степени обработанный сигнал соответствует входному сигналу или идеальному сигналу. При отсутствии разницы система является полностью прозрачной. Второй тип описывает, в какой степени обработанный сигнал соответствует входному сигналу или идеальному сигналу для человека-наблюдателя. Если в экспериментальных условиях какая-либо разница не может быть заметна, система является прозрачной для восприятия. Термин "прозрачный" при отсутствии эксплицитной ссылки на критерий будет использоваться для систем, которые являются прозрачными для восприятия.

3.7 Дублирование (replication) – повторяющееся представление одних и тех же условий цепи (с тем же исходным материалом) тому же самому субъекту.

3.8 Надежность субъективного теста (reliability of a subjective test):

- a) надежность личного теста (с участием одного лица) означает совпадение повторяющихся оценок какого-либо субъекта в отношении одинаковых условий теста;
- b) надежность совместного теста (с участием нескольких лиц) означает совпадение оценок различных субъектов в отношении одинаковых условий теста.

3.9 Достоверность субъективного испытания (validity of a subjective test) – соответствие между средним значением показателей, полученных в ходе испытания, и истинным значением, для измерения которого предназначено испытание.

3.10 Эталонные условия (reference conditions) – условия с использованием манекена, добавленные к условиям испытания с целью зафиксировать оценки, полученные в ходе различных экспериментов.

3.11 Прямая ссылка (исходная ссылка) (explicit reference (source reference)) – условие, используемое экспертами в качестве ссылки для выражения своего мнения, когда используется метод DCR. Эта ссылка представляется в начале в каждой паре последовательностей. Обычно форматом прямой ссылки является формат, используемый на входе тестируемых кодеков (например, [ITU-R BT.601-4], CIF, QCIF, SIF и т. д.). В тексте настоящей Рекомендации, там где значение "ссылка" явствует из контекста, слова "прямая" и "исходная" опускаются.

3.12 Косвенная ссылка (implicit reference) – условие, используемое экспертами в качестве ссылки для выражения своего мнения о тестируемом материале, когда используется метод АСР. Если косвенная ссылка предлагается экспериментатором, она должна быть хорошо известна всем экспертам (например, стандартные ТВ-системы, реальные условия).

4 Сокращения и акронимы

В настоящей Рекомендации используются следующие сокращения и акронимы.

ACR	Absolute Category Rating		Показатель абсолютной категории
ACR-HR	Absolute Category Rating with Hidden Reference		Показатель абсолютной категории со скрытым эталоном
CCD	Charge Coupled Device	ПЗС	Прибор с зарядовой связью
CI	Confidence Interval		Доверительный интервал
CIF	Common Intermediate Format		Единый промежуточный формат

ПРИМЕЧАНИЕ. – Формат изображения, определенный в [b-ITU-T H.261] для видеотелефона: 352 строки × 288 пикселей.

CRT	Cathode Ray Tube	ЭЛТ	Электронно-лучевая трубка
DCR	Degradation Category Rating		Показатель ухудшения категории качества
DV	Differential viewer		Дифференциальная зрительская оценка
%GOB	Percent of Good or Better (proportion of Good and Excellent)		Процент оценок "хорошо" или "лучше" (доля голосов "хорошо" и "отлично")
LCD	Liquid Crystal Display	ЖКД	Жидкокристаллический дисплей
MOS	Mean Opinion Score		Средняя экспертная оценка
PC	Pair Comparison		Сравнение пар
%POW	Percent of Poor or Worse (proportion of Poor and Bad votes)		Процент оценок "плохо" или "хуже" (доля голосов "плохо" и "неудовлетворительно")
PVS	Processed Video Sequence		Обработанная последовательность видеок кадров
QCIF	Quarter CIF		Четверть CIF

ПРИМЕЧАНИЕ. – Формат изображения, определенный в [b-ITU-T H.261] для видеотелефона: 176 строк × 144 пикселя.

S/N	Signal-to-Noise ratio		Отношение сигнал/шум
SI	Spatial Information		Пространственная информация
SIF	Standard Intermediate Format		Стандартный промежуточный формат

ПРИМЕЧАНИЕ. – Формат изображения, определенный в [b-ISO/IEC 11172] (MPEG-1): 352 строки × 288 пикселей × 25 кадров/с и 352 строки × 240 пикселей × 30 кадров/с.

SP	Simultaneous Presentation		Одновременное представление
std	Standard Deviation		Стандартное отклонение
TI	Temporal Information		Временная информация
VTR	Video Tape Recorder		Видеомагнитофон

5 Исходный сигнал

Для управления характеристиками исходного сигнала должны быть определены и записаны в цифровой системе хранения тестовые последовательности в соответствии с целью испытаний. Если испытателя интересует сравнение результатов, полученных в различных лабораториях, необходимо использовать общий набор исходных последовательностей для исключения дополнительного источника разброса.

5.1 Условия записи

Источники света (лампы накаливания или люминесцентные лампы) могут располагаться над камерой или рядом с ней. При размещении источников света следует учитывать, что для офисного освещения более типично верхнее расположение, и именно оно должно использоваться в сценах, изображающих деловую среду. Следует избегать студийных ламп и других нестандартных источников света.

В помещениях условия освещения в поле зрения могут варьироваться от 100 люкс до примерно 10 000 люкс. Необходимо учитывать колебания (с частотой переменного тока) света (при флуоресцентном освещении), поскольку это может вызвать мерцание в записи.

Следует тщательно контролировать и указывать условия освещения, цвет стен, отражательную способность поверхностей и т. д.

5.2 Система записи

5.2.1 Камера

Последовательности видеоизображений следует записывать с помощью высококачественной ПЗС-камеры.

Отношение сигнал/шум входного видеосигнала может оказывать значительное влияние на характеристики кодека.

Для определения входного видеосигнала необходимо указать следующие параметры:

- динамический диапазон сигналов YUV;
- коэффициент гамма-коррекции (должен составлять 0,45);
- ширину полосы пропускания/крутизну характеристики фильтров;
- чувствительность камеры при очень низких уровнях освещения и характеристики автоматического регулирования усиления (APU), если оно используется.

Взвешенное отношение S/N следует измерять в соответствии с [ITU-T J.61], часть С пункт 3.2.1. Среднеквадратическое значение взвешенного отношения S/N должно превышать 45 дБ.

Нестабильность или дрожание тактовых сигналов может вызвать шумовые эффекты. Для устройства синхронизации камеры требуется стабильность не менее 0,5 част./млн.

Можно использовать системы с фиксированным или переменным фокусным расстоянием. Для настольных терминалов разумная глубина фокусировки составляет от 30 см до 120 см, а для многопользовательских систем более подходящей считается глубина фокусировки от 50 см до бесконечности. Чтобы учесть изменение освещенности в помещении для записи, следует использовать регулируемую диафрагму или нейтральные светофильтры. Камера должна иметь функцию автоматического баланса белого, чтобы адаптироваться к цветовой температуре источника света. Коррекция температуры белого цвета может варьироваться от 2700 °К (в помещении с лампой накаливания) до 6500 °К (температура рассеянного дневного света).

5.2.2 Формат видеосигнала и его хранения

Исходные видеосигналы, обеспечиваемые видеокамерой, должны быть дискретизированы в соответствии с [ITU-R BT.601-4], часть А. Во избежание искажений исходного сигнала его следует хранить в цифровом формате, например в компьютере или на магнитной ленте формата D1 4 : 2 : 2.

5.3 Характеристики сцены

Один из важных вопросов – это выбор тестовых сцен. В частности, критическими параметрами является информация о пространственном и информация о временном восприятии сцен. Эти параметры играют решающую роль в определении возможной величины сжатия видеоизображения и, следовательно, уровня искажения при передаче сцены по цифровому служебному каналу передачи с фиксированной скоростью. Для тестовой видеозаписи следует выбирать достоверные и релевантные сцены, чтобы их пространственная и временная информация соответствовала тем видеослугам, которые должны передаваться по цифровому служебному каналу. Набор тестовых сцен должен охватывать весь спектр пространственной и временной информации, представляющей интерес для пользователей тестируемых устройств.

Детали, связанные с характеристикой тестовых последовательностей, и примеры подходящих тестовых сцен приведены в Приложении А и Дополнениях I и II.

Количество последовательностей должно определяться в соответствии с планом эксперимента. Во избежание утомления наблюдателей и в целях достижения минимальной надежности результатов для последовательностей необходимо выбрать сцены по крайней мере четырех разных типов (то есть на разные темы).

В следующих подпунктах представлены методы количественной оценки пространственной и временной информации тестовых сцен. Эти методы применимы к тестированию качества видеоизображений как в настоящее время, так и в будущем. Важную роль играет расположение видеосцены в пространственно-временной матрице, поскольку от этого часто в значительной степени зависит качество передаваемой видеосцены (особенно после прохождения через кодек с низкой битовой скоростью). Представленная здесь количественная оценка пространственной и временной информации может использоваться для обеспечения надлежащего охвата пространственно-временной плоскости.

Представленные ниже измерения пространственной и временной информации имеют единственное значение для каждого кадра полной тестовой последовательности. В результате получается временной ряд значений, обычно различающихся в некоторой степени. Представленные ниже измерения информации о восприятии устраняют эту изменчивость с помощью максимальной функции (максимальное значение для последовательности). Сама изменчивость может быть с успехом изучена, например, с помощью графиков пространственно-временной информации, построенных на поккадровой основе. Использование распределений информации по тестовой последовательности позволяет также лучше оценить сцены, содержащие смонтированные сцены.

5.3.1 Измерение информации о пространственном восприятии

Информация о пространственном восприятии (SI) основана на фильтре Собеля. Каждый кадр видеоизображения (плоскость яркости) в момент времени n (F_n) сначала фильтруется фильтром Собеля [Sobel (F_n)]. Затем рассчитывается стандартное отклонение по пикселям (std_{space}) в каждом кадре, прошедшем через фильтр Собеля. Данная операция повторяется для каждого кадра последовательности видеоизображения, и в результате получается временной ряд пространственной информации сцены. Для представления содержания пространственной информации сцены из временного ряда выбирается максимальное значение (max_{time}). Этот процесс может быть представлен в форме уравнения:

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

5.3.2 Измерение информации о временном восприятии

Информация о временном восприятии (ТІ) основана на функции разницы, вызванной движением $M_n(i, j)$, которая представляет собой разность между значениями пикселей (в плоскости яркости) в одном и том же месте в пространстве, но в последовательные моменты времени или в последовательных кадрах. $M_n(i, j)$ как функция времени (n) определяется следующим образом:

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

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

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

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

Более резкое движение в смежных кадрах приводит к более высоким значениям ТІ.

ПРИМЕЧАНИЕ. – Для сцен, содержащих переходы от сцены к сцене, могут быть заданы два значения – одно с учетом перехода при измерении временной информации, а другое без него.

6 Методы испытаний и планирование эксперимента

Для измерения воспринимаемого качества изображений требуются методы субъективного шкалирования. Условием значимости таких измерений является наличие связи между физическими характеристиками источника воздействия (стимула), в данном случае последовательности видеок кадров (видеоряда), представленной участникам эксперимента, и силой и характером ощущений, вызванных этим воздействием.

Известен целый ряд проверенных экспериментальных методов для разных целей. Здесь рекомендуются три метода для приложений, использующих каналы передачи данных с битовыми скоростями, определенными в классах TV3, MM4, MM5 и MM6, как указано в таблице D.2. Другие методы испытаний описаны в Дополнениях III и IV.

Окончательный выбор одного из этих методов для конкретного приложения зависит от нескольких факторов, таких как контекст, назначение и стадия процесса разработки, на которой проводится испытание.

6.1 Показатель абсолютной категории (ACR)

Метод оценки абсолютной категории – это метод оценки по категориям, когда каждая тестовая последовательность демонстрируется отдельно и оценивается по шкале категорий независимо. (Этот метод также называется методом с одним источником воздействия.)

Согласно этому методу после каждой демонстрации наблюдателям предлагается оценить качество показанной последовательности.

Порядок демонстрации источников воздействия во времени показан на рисунке 1. Если используется постоянное время голосования (например, несколько зрителей работают с видеозаписью одновременно), то оно не должно превышать 10 с. Время демонстрации можно уменьшить или увеличить в зависимости от содержания тестового материала.

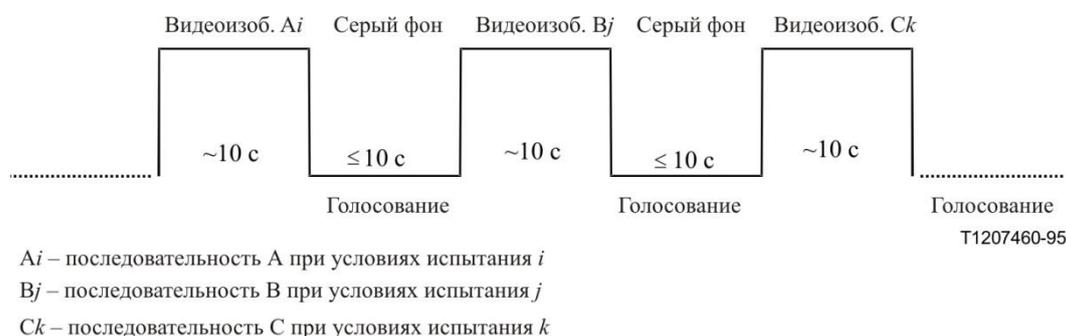


Рисунок 1 – Представление источников воздействия при использовании метода ACR

Для оценки общего качества должна использоваться следующая пятибалльная шкала:

- 5 отлично;
- 4 хорошо;
- 3 удовлетворительно;
- 2 неудовлетворительно;
- 1 плохо.

Если требуется более высокая дискриминирующая способность, можно использовать девятибалльную шкалу. Примеры подходящих числовых и непрерывных шкал приведены в Приложении В. Там же даны примеры рейтинговых оценок, дополняющих оценку общего качества. Такие оценки могут быть полезны для получения дополнительной информации о различных факторах качества восприятия, когда для некоторых тестируемых систем общие оценки качества почти одинаковы, хотя системы явно воспринимаются по-разному.

Необходимое количество повторений для метода ACR получают путем воспроизведения одних и тех же условий в разные моменты времени в ходе испытаний.

6.2 Показатель абсолютной категории со скрытым эталоном (ACR-HR)

Метод оценки абсолютной категории со скрытым эталоном – это метод оценки по категориям, когда каждая тестовая последовательность демонстрируется отдельно и оценивается по шкале категорий независимо. Имеющаяся процедура испытаний должна включать эталонную версию каждой тестовой последовательности, которая демонстрируется как любой другой тестовый источник воздействия. Это называется скрытым эталонным условием. Во время анализа данных вычисляется дифференциальная оценка качества (DMOS) между каждой тестовой последовательностью и соответствующим (скрытым) эталоном. Такая процедура называется "скрытый эталон".

Согласно этому методу после каждой демонстрации наблюдателям предлагается оценить качество показанной последовательности.

Порядок демонстрации источников воздействия во времени показан на рисунке 1. Если используется постоянное время голосования (например, несколько зрителей работают с видеозаписью одновременно), то оно не должно превышать 10 с. Время демонстрации можно уменьшить или увеличить в зависимости от содержания тестового материала.

Для оценки общего качества должна использоваться следующая пятибалльная шкала:

- 5 отлично;
- 4 хорошо;
- 3 удовлетворительно;
- 2 неудовлетворительно;
- 1 плохо.

По каждому наблюдателю на основе обработанной последовательности видеок кадров (PVS) рассчитываются дифференциальные зрительские оценки (DV). DV вычисляется с использованием соответствующего скрытого эталона (REF) по следующей формуле:

$$DV(PVS) = V(PVS) - V(REF) + 5,$$

где V – зрительская оценка ACR. При использовании этой формулы $DV = 5$ означает "отличное", а $DV = 1$ – "плохое" качество. Любые значения DV больше 5 (то есть когда обработанная последовательность получает лучшую оценку, чем соответствующий скрытый эталон), как правило, считаются действительными. В качестве альтернативы можно применять функцию сжатия по двум точкам для предотвращения нежелательного влияния этих индивидуальных зрительских оценок (DV) ACR-HR на общую среднюю экспертную оценку:

$$\text{crushed_DV} = (7 * DV) / (2 + DV), \text{ когда } DV > 5.$$

Если требуется более высокая дискриминирующая способность, можно использовать девятибалльную шкалу ACR. Примеры подходящих числовых и непрерывных шкал приведены в Приложении В. Там же приведены примеры рейтинговых оценок, дополняющих оценки общего качества. Такие оценки могут быть полезны для получения дополнительной информации о различных факторах качества восприятия, когда для некоторых тестируемых систем общие оценки качества почти одинаковы, хотя системы явно воспринимаются по-разному.

Необходимое количество повторений для метода ACR-HR получают путем воспроизведения одних и тех же условий в разные моменты времени в ходе испытаний.

Метод ACR-HR должен использоваться только с тем эталонным видеозображением, которому эксперты в этой области дают оценку "хорошо" или "отлично" по вышеупомянутой пятибалльной шкале.

Метод ACR-HR не подходит для анализа необычных нарушений, происходящих на первой и последней секундах видеоряда. Незнакомство зрителя с эталонной последовательностью видеокadra может привести к пропуску явного нарушения (например, если видеоряд приостанавливается непосредственно перед концом, зритель не сможет определить, относится ли это к его содержанию или к ошибке сети).

6.3 Оценка ухудшения категории качества (DCR)

Оценка ухудшения категории качества предполагает, что тестовые последовательности представляются парами: первый источник воздействия в каждой паре всегда является исходным эталоном, а второй – это та же исходная последовательность, демонстрируемая через одну из тестируемых систем. (Этот метод также называется методом шкалы ухудшения с двумя источниками воздействия.)

При использовании уменьшенных форматов изображения (например, CIF, QCIF, SIF) может оказаться полезной одновременная демонстрация эталонной и тестовой последовательностей на одном и том же мониторе. Руководящие указания по такой процедуре представления рассматриваются в Приложении С.

Порядок демонстрации источников воздействия во времени показан на рисунке 2. Если используется постоянное время голосования (например, несколько зрителей работают с видеозаписью одновременно), то оно не должно превышать 10 с. Время демонстрации можно уменьшить или увеличить в зависимости от содержания тестового материала.

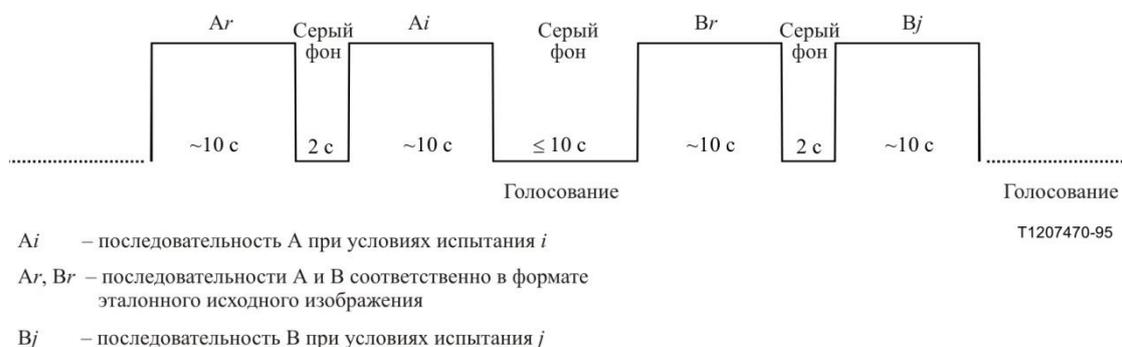


Рисунок 2 – Представление источников воздействия при использовании метода DCR

В этом случае наблюдателям предлагается оценить ухудшение качества второго источника воздействия относительно эталона.

Для оценки ухудшения должна использоваться следующая пятибалльная шкала:

- 5 незаметное;
- 4 заметное, но не раздражающее;
- 3 немного раздражающее;
- 2 раздражающее;
- 1 сильно раздражающее.

Необходимое количество повторений для метода DCR получают путем воспроизведения одних и тех же условий в разные моменты времени в ходе испытаний.

6.4 Метод парного сравнения (РС)

Метод парного сравнения подразумевает, что тестовые последовательности представляются парами, состоящими из одной и той же последовательности, которая демонстрируется сначала через одну тестируемую систему, а затем через другую.

Тестируемые системы (А, В, С и т. д.) обычно объединяются в $n(n - 1)$ всех возможных комбинаций – АВ, ВА, СА и т. д. Таким образом все пары последовательности должны демонстрироваться в двух возможных порядках (например, АВ и ВА). После демонстрации каждой пары выносится суждение о том, какой из элементов пары является предпочтительным в контексте сценария испытаний.

Порядок демонстрации источников воздействия во времени показан на рисунке 3. Если используется постоянное время голосования (например, несколько зрителей работают с видеозаписью одновременно), то оно не должно превышать 10 с. Время демонстрации должно составлять около 10 с, и его можно уменьшить или увеличить в зависимости от содержания тестируемого материала.

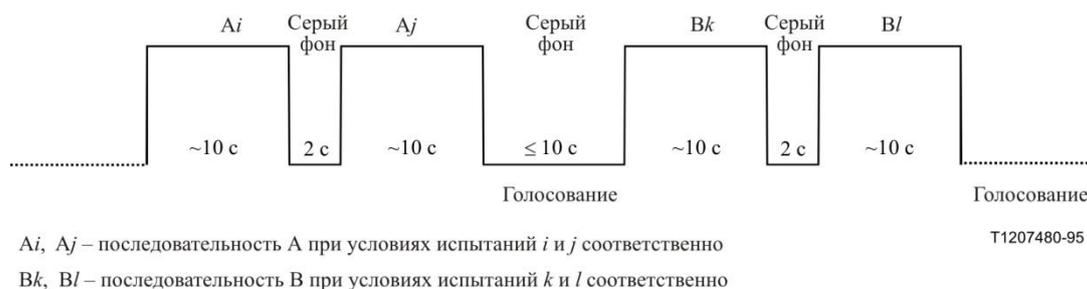


Рисунок 3 – Представление источников воздействия при использовании метода РС

При использовании уменьшенного разрешения (например, CIF, QCIF, SIF) может оказаться полезной одновременная демонстрация каждой пары последовательностей на одном и том же мониторе. Руководящие указания по такой процедуре представления рассматриваются в Приложении С.

В случае применения метода РС количество повторений обычно не рассматривается, поскольку сам метод предполагает повторное представление одних и тех же условий, хотя и в разных парах.

Существует разновидность метода РС с использованием категориальной шкалы для дальнейшего измерения различий между парой последовательностей. См. [ITU-R BT.500-9] и [ITU-T P.800].

6.5 Сравнение методов

Важной проблемой при выборе метода тестирования является фундаментальное различие между методами с использованием явных эталонов (explicit reference) (например, DCR) и методами, при которых явные эталоны не используются (например, ACR, ACR-HR и РС). Методы этой второй категории не проверяют прозрачность или точность передачи.

Метод DCR следует использовать при проверке точности передачи по отношению к исходному сигналу. Часто это является одним из важных факторов при оценке систем высокого качества. DCR давно стал ключевым методом оценки телевизионных изображений, типичное качество которых соответствует чрезвычайно высокому уровню качества видеотелефонии и видеоконференц-связи, и описан в [ITU-R BT.500-9]. Для оценки систем высокого качества могут использоваться и другие методы. Когда важным фактором является наблюдение ухудшения зрителем, ценной информацией служат уточняющие комментарии шкалы DCR (незаметное/заметное).

Таким образом, когда важно проверить точность передачи в отношении исходного сигнала, следует использовать метод DCR.

Метод DCR также следует применять для оценки высококачественных систем в контексте мультимедийной связи. Этому способствует разделение незаметного/заметного ухудшения на шкале DCR, а также сравнение с эталонным качеством.

ACR является легко и быстро реализуемым методом, а представление источников воздействия аналогично их представлению при обычном использовании систем. Поэтому ACR хорошо подходит для квалификационных испытаний.

Метод ACR-HR обладает всеми достоинствами метода ACR в отношении представления и скорости. Главное преимущество ACR-HR перед ACR заключается в том, что из субъективных оценок можно исключить перцептивное воздействие эталонного видеоизображения. Это уменьшает влияние на окончательные оценки смещения сцены (например, нравится или не нравится зрителям эталонное видеоизображение), качества эталонного видеоизображения (например, небольшие различия в качестве камеры) и монитора (например, монитора профессионального качества по сравнению с потребительским). ACR-HR хорошо подходит для крупных экспериментов при условии, что все эталонные видеоизображения имеют по крайней мере "хорошее" качество. Однако ACR-HR может быть нечувствительным к некоторым ухудшениям качества, которые легко обнаруживаются с помощью методов прямого сравнения (например, DCR). Так, методом ACR-HR нельзя обнаружить систематическое снижение коэффициента усиления цвета (например, притушенные цвета).

Главным достоинством метода РС является его высокая дискриминационная способность, которая имеет особое значение при тестировании нескольких образцов почти эквивалентного качества.

Когда в ходе одного и того же испытания нужно оценить большое количество образцов, процедура, основанная на методе РС, получается длинной. В этом случае сначала можно провести тестирование с помощью метода ACR или DCR с ограниченным числом наблюдателей, а затем тестирование с использованием метода РС только для тех образцов, которые получили примерно одинаковые оценки.

6.6 Эталонные условия

Результаты оценки качества часто зависят не только от фактического качества видеоизображения, но и от других факторов, таких как общий диапазон качества условий испытаний, опыт и ожидания оценщиков и т. д. Чтобы контролировать некоторые из этих эффектов, можно добавить ряд фиктивных условий испытаний и использовать их в качестве эталона.

Описание эталонных условий и процедур их получения приведено в [ITU-T P.930]. Когда ухудшения, вносимые тестовыми элементами, невелики, при тестировании с использованием метода РС особенно рекомендуется ввести в качестве эталонного условия исходный сигнал.

Уровень качества эталонных условий должен охватывать по крайней мере диапазон качества тестовых элементов.

6.7 Планирование эксперимента

Могут использоваться разные планы экспериментов, такие как полностью рандомизированный план, латинский и греко-латинский квадраты, квадрат Юдена, повторяющиеся блоки и т. д. [b-Kirk], выбор которых должен зависеть от цели эксперимента.

Метод планирования выбирается экспериментатором самостоятельно, с тем чтобы соответствовать конкретным целям по затратам и точности. План может зависеть от того, какие условия представляют особый интерес для данного теста.

Рекомендуется включить в эксперимент по крайней мере две, а по возможности три или четыре реплики (то есть повторения идентичных условий). Существует несколько причин для применения реплик, наиболее важной из которых является то, что с использованием реплицированных данных можно производить измерения "внутрисубъектной вариации". Для проверки надежности участника (субъекта) можно использовать один и тот же порядок представления в одинаковых условиях. Если используется другой порядок представления, то результирующая вариация экспериментальных данных состоит из эффекта порядка и внутрисубъектной вариации.

Реплики позволяют рассчитать надежность отдельных результатов и при необходимости исключить ненадежные результаты, полученные от некоторых участников. Кроме того, оценка внутрисубъектного и межсубъектного стандартного отклонения является предпосылкой для правильного анализа расхождений и для обобщения результатов на более широкую аудиторию. В дополнение к этому в какой-то мере компенсируются эффекты, связанные с обучением в ходе испытаний.

Еще большее снижение влияния эффектов, связанных с обучением, достигается путем включения нескольких учебных демонстраций в начале каждого сеанса испытаний. Они должны быть подобны демонстрациям, которые будут показаны позднее в ходе испытания. Предварительные демонстрации не должны приниматься во внимание при статистическом анализе результатов испытаний.

7 Процедуры оценки

В таблице 1 приведены типичные условия просмотра, используемые при оценке качества видеоизображения. Должны быть указаны фактические установленные значения (установки) параметров, используемые при оценке. Для сравнения результатов все условия просмотра в лабораториях, проводящих испытания одного и того же типа, должны быть фиксированными и равными.

Размер и тип используемого устройства отображения должны подходить для испытываемого приложения. Когда последовательности демонстрируются через систему на базе ПК, должны быть указаны характеристики дисплея, например шаг триад монитора, тип используемой видеокарты и т. д.

Что касается формата отображения, то для отображения последовательностей предпочтительно использовать весь экран. Тем не менее когда по какой-либо причине последовательности должны отображаться в окне экрана, цвет фона на экране должен быть 50%-ным серым, соответствующим $Y = U = V = 128$ (U и V без знака).

7.1 Условия просмотра

Испытания должны проводиться при следующих условиях просмотра.

Таблица 1 – Условия просмотра

Параметр	Установка
Расстояние просмотра (Примечание 1)	1–8H (Примечание 2)
Пиковая яркость экрана	100–200 кд/м (Примечание 2)
Отношение яркости темного экрана к пиковой яркости	≤ 0,05
Отношение яркости экрана при отображении только уровня черного в совершенно темной комнате к уровню, соответствующему пиковому белому	≤ 0,1
Отношение яркости фона за монитором к пиковой яркости изображения (Примечание 3)	≤ 0,2
Цветность фона (Примечание 4)	D ₆₅
Фоновая освещенность помещения (Примечание 3)	≤ 20 люкс
<p>ПРИМЕЧАНИЕ 1. – Вероятно, что при данной высоте экрана, когда качество изображения ухудшается, расстояние просмотра, предпочитаемое наблюдателями, будет увеличиваться. Что касается этого параметра, для квалификационных испытаний предпочтительное расстояние просмотра должно быть определено заранее. Расстояние просмотра обычно зависит от приложения.</p> <p>ПРИМЕЧАНИЕ 2. – H – высота изображения. Расстояние просмотра должно определяться с учетом не только размера, но и типа экрана, типа приложения и цели эксперимента.</p> <p>ПРИМЕЧАНИЕ 3. – Данное значение указывает установку, при которой обеспечивается максимальная возможность обнаружения искажений. Для некоторых приложений допускаются большие значения или они определяются приложением.</p> <p>ПРИМЕЧАНИЕ 4. – Для ПК-мониторов цвет фона может быть адаптирован к цветности монитора.</p>	

7.2 Система обработки и воспроизведения

Существует два метода получения тестовых изображений из исходных записей:

- путем передачи или воспроизведения видеозаписей в режиме реального времени через тестируемую систему, в то время как наблюдатели просматривают их и реагируют на них;
- путем автономной обработки исходных записей через тестируемое устройство и записи результата с созданием нового набора записей.

Во втором случае следует использовать цифровой видеомэгнитофон для минимизации ухудшений качества, которые могут быть вызваны процессом записи. В любом случае могут использоваться профессиональные видеомэгнитофоны, такие как D2, MII и BetacamSP, с учетом того, что искажения, введенные схемами кодирования с низкой битовой скоростью, обычно более заметны, чем искажения, вносимые модуляцией.

Можно использовать ЭЛТ-, ЖК-, плазменный, проекционный или другой монитор с учетом типа приложения и цели эксперимента. Размер и тип используемого монитора должны подходить для исследуемого приложения.

Мониторы должны быть откалиброваны в соответствии с процедурами, определенными в [ITU-R BT.814-1].

7.3 Зрители

Возможное количество участников тестового просмотра (а также тестирования удобства использования терминалов или услуг) составляет от 4 до 40. Четыре участника – это абсолютный минимум по статистическим соображениям, а за пределы 40 это число выходит редко.

На самом деле фактическое число участников конкретного испытания зависит от требуемой достоверности и необходимости обобщения результатов выборки на широкую аудиторию.

Как правило, в эксперименте должны участвовать не менее 15 наблюдателей. Они не должны быть опытными оценщиками, и их работа не должна быть непосредственно связана с оценкой качества изображений.

Тем не менее на ранних этапах разработки систем видеосвязи и в предварительных экспериментах, проводимых перед крупными испытаниями, ориентировочные результаты могут дать небольшие группы экспертов (4–8 человек) или других важных субъектов.

Перед сеансом наблюдателей обычно проверяют на предмет нормальной остроты зрения или коррекции до нормальной остроты и на нормальное цветовосприятие. Что касается остроты зрения, то у наблюдателя не должно быть ошибок на строке стандартной оптометрической таблицы, соответствующей остроте зрения 20/30 [b-Snellen]. Таблица масштабируется для расстояния тестового просмотра, и острота зрения проверяется в том же месте, где будут просматриваться видеоизображения (то есть оптометрическая таблица вешается на монитор) при сидящих зрителях. Что касается цветовосприятия, то допускается не более двух ошибок из 12 карточек [b-Beck].

7.4 Инструктаж зрителей и учебный сеанс

Перед началом эксперимента зрителям следует предоставить сценарий предполагаемого применения тестируемой системы. Кроме того, им в письменной форме предоставляется описание типа оценки, шкалы экспертной оценки и представления источников воздействия. Диапазон и тип ухудшений демонстрируются в ходе предварительных испытаний, которые могут содержать последовательности видеоклипов, отличные от тех, которые используются в реальных тестах.

Наихудшему качеству, продемонстрированному в обучающем наборе, необязательно должна соответствовать самая низкая субъективная оценка по шкале.

Во избежание предвзятости отвечать на вопросы о процедуре или о значении инструкций следует с осторожностью и только до начала сеанса.

Возможный текст инструкций для оценщиков приведен в Дополнении II.

8 Статистический анализ и представление результатов

Результаты должны представляться вместе с подробным описанием экспериментальной установки. Для каждой комбинации тестовых переменных следует указывать среднее значение и стандартное отклонение статистического распределения оценок.

По этим данным рассчитывается надежность участников и указывается метод, используемый для ее оценки. Некоторые критерии определения надежности участников приведены в [ITU-R BT.500-9] и [IEC/TR 60268-13].

Полезно проанализировать кумулятивное распределение баллов. Поскольку кумулятивные распределения нечувствительны к линейности, они могут быть особенно полезными для данных с сомнительной линейностью, таких как данные, полученные с использованием методов ACR и DCR, вместе с категорийной шкалой без градаций (то есть оценкой по категориям).

Данные могут быть организованы, например, как показано в таблице 2 для ACR.

Таблица 2 – Информативная таблица с кумулятивным распределением баллов для метода ACR

Условие	Всего голосов	Отлично	Хорошо	Удовлетворительно	Неудовлетворительно	Плохо	MOS	CI	Std	%GOB	%POW

Условие – обозначение комбинации тестовых переменных.
Всего голосов – число голосов, полученных для этого условия.
Отлично, хорошо, ... плохо – проявление каждого голоса.

Для оценки значимости параметров испытаний должны использоваться классические методы анализа расхождений. Если целью является оценка качества видеоизображения в зависимости от параметра, то для интерпретации данных могут быть полезны методы подбора аппроксимирующей кривой.

Для случая парных сравнений в Справочнике по телефонometriи [b-ITU-T Handbook], раздел 2.6.2C, описан метод вычисления положения каждого источника воздействия на шкале интервалов, где разница между такими источниками соответствует разнице в предпочтениях.

Приложение А

Детали, связанные с характеристикой тестовых последовательностей

(Данное Приложение является неотъемлемой частью настоящей Рекомендации)

А.1 Фильтр Собеля

Фильтр Собеля реализуется путем свертывания двух ядер 3×3 по видеокадру и извлечения квадратного корня из суммы квадратов результатов этих сверток.

Для $y = \text{Sobel}(x)$, пусть $x(i, j)$ – пиксель входного изображения в i -й строке и j -м столбце. Результат первой свертки $Gv(i, j)$ равен:

$$\begin{aligned} Gv(i, j) = & -1 \times x(i-1, j-1) - 2 \times x(i-1, j) - 1 \times x(i-1, j+1) + \\ & + 0 \times x(i, j-1) + 0 \times x(i, j) + 0 \times x(i, j+1) + \\ & + 1 \times x(i+1, j-1) + 2 \times x(i+1, j) + 1 \times x(i+1, j+1). \end{aligned}$$

Аналогично результат второй свертки $Gh(i, j)$ равен:

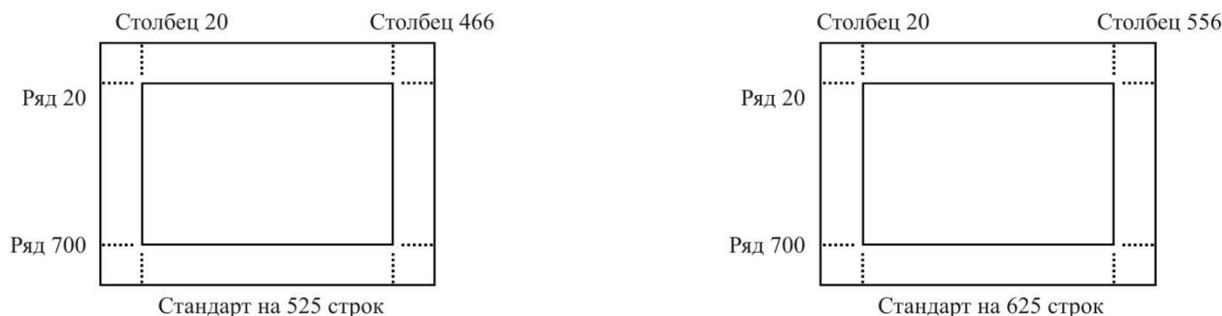
$$\begin{aligned} Gh(i, j) = & -1 \times x(i-1, j-1) + 0 \times x(i-1, j) + 1 \times x(i-1, j+1) + \\ & - 2 \times x(i, j-1) + 0 \times x(i, j) + 2 \times x(i, j+1) + \\ & - 1 \times x(i+1, j-1) + 0 \times x(i+1, j) + 1 \times x(i+1, j+1). \end{aligned}$$

Следовательно, результат обработки изображения фильтром Собеля в i -й строке и j -м столбце равен:

$$y(i, j) = \sqrt{[Gv(i, j)]^2 + [Gh(i, j)]^2}.$$

Расчеты выполняются для всех $2 \leq i \leq N-1$ и $2 \leq j \leq M-1$, где n – число строк, а M – число столбцов.

Рекомендуется выполнять расчеты на фрагменте видеокадра во избежание нежелательных краевых эффектов, а также потому, что пользователям ЭЛТ края видеокадра обычно невидимы. Это можно сделать с использованием подходящего фрагмента, например, как показано на рисунке А.1 для форматов 625 и 525 строк [ITU-R BT.601-4].



T1207490-95

Рисунок А.1 – Фрагменты кадра, используемые для расчета значений SI и TI для форматов 525 и 625 строк [ITU-R BT.601-4]

Подробнее о фильтре Собеля см. в [b-Gonzalez].

А.2 Как использовать SI и TI для выбора тестовой последовательности

При выборе тестовых последовательностей бывает полезно сравнить относительную пространственную и временную информацию разных доступных последовательностей. Как правило, трудность сжатия напрямую связана с пространственной и временной информацией последовательности.

Если в данном тесте должно использоваться небольшое количество тестовых последовательностей, важно выбрать такие, которые охватывают большую часть пространственно-временной информационной плоскости (см. рисунок А.2). При использовании четырех тестовых последовательностей можно выбрать последовательность из каждого из четырех квадрантов пространственно-временной информационной плоскости.

Если же нужно выбрать тестовые последовательности, эквивалентные по трудности кодирования, то желательно выбрать последовательности с подобными значениями SI и TI.

А.3 Примеры

На рисунке А.2 показаны относительное количество пространственной и временной информации для некоторых репрезентативных тестовых сцен и способы их размещения на пространственно-временной информационной плоскости.

Вдоль оси $TI = 0$ (в нижней части диаграммы) показаны неподвижные сцены и сцены с очень ограниченным движением (такие как *l*, *f* и *a*). В верхней части диаграммы находятся сцены с интенсивным движением (такие как *p*, *q* и *i*). Вдоль оси $SI = 0$ (по левому краю диаграммы) расположены сцены с минимумом пространственной информации (такие как *l*, *k*, *x*, *u* и *f*). Рядом с правым краем диаграммы расположены сцены с максимальным количеством пространственной информации (такие как *h* и *s*). Значения SI и TI получены с использованием приведенных выше уравнений и видеоизображения с пространственным шаблоном в соответствии со спецификациями [ITU-R ВТ.601-4]. В таблице А.1 приведены примеры тестовых сцен в разбивке по категориям содержания сцены.

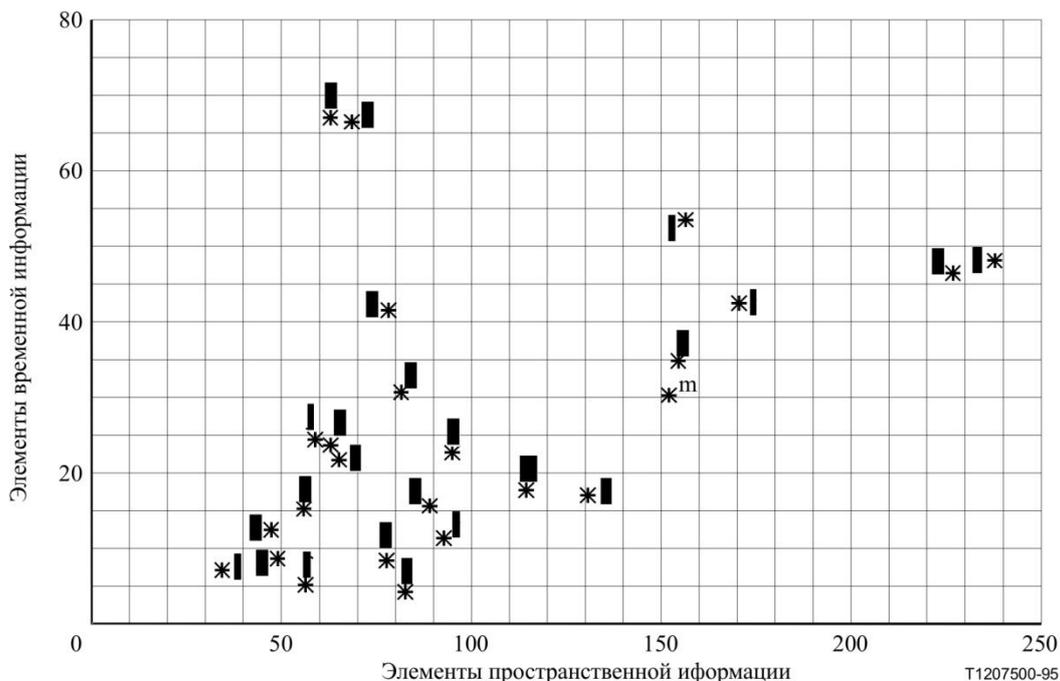


Рисунок А.2 – Пространственно-временная диаграмма примерного набора тестовых сцен

Таблица А.1 – Категории содержания сцены

Категория	Описание	Наименование и буквенное обозначение сцены
<i>A</i>	Один человек, в основном голова и плечи, ограниченная детализация и подвижность	vtc1nw(f), susie(j), disguy(k), disgal(l)
<i>B</i>	Один человек с графикой и/или большей детализацией	vtc2mp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w), inspec(x)
<i>C</i>	Несколько человек	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
<i>D</i>	Графика с указанием	washdc(c), cirkit(s), rodmap(t), filter(u), ysmite(v),
<i>E</i>	Интенсивное движение объекта и/или камеры (примеры широкоэкранный ТВ)	flogar(h), ftball(i), fedas(y)

Приложение В

Дополнительные оценочные шкалы

(Данное Приложение является неотъемлемой частью настоящей Рекомендации)

В.1 Рейтинговые шкалы

Часто бывает необходимо использовать рейтинговую шкалу с числом градаций больше пяти, в частности для оценки видеокодеков с низкой битовой скоростью. Для этой цели подходит девятибалльная шкала, в которой пять словесно определяемых категорий качества, как рекомендовано в пункте 6.1, используются в качестве ярлыков для каждой второй оценки шкалы, как показано на рисунке В.1.

9	Отлично
8	
7	Хорошо
6	
5	Удовлетворительно
4	
3	Неудовлетворительно
2	
1	Плохо

Рисунок В.1 – Девятибалльная числовая шкала качества

Дальнейшее расширение этой шкалы показано на рисунке В.2, где конечные точки словесно определены как точки привязки, не используемые для оценки. В этом словесном определении используется некоторый эталон (например, на рисунке В.2 в качестве эталона используется оригинал). Этот эталон может быть явным или неявным и четко иллюстрируется на этапе обучения. См. также [IEC/TR 60268-13] и [b-ITU-T Handbook], раздел 2.6, шкала а).

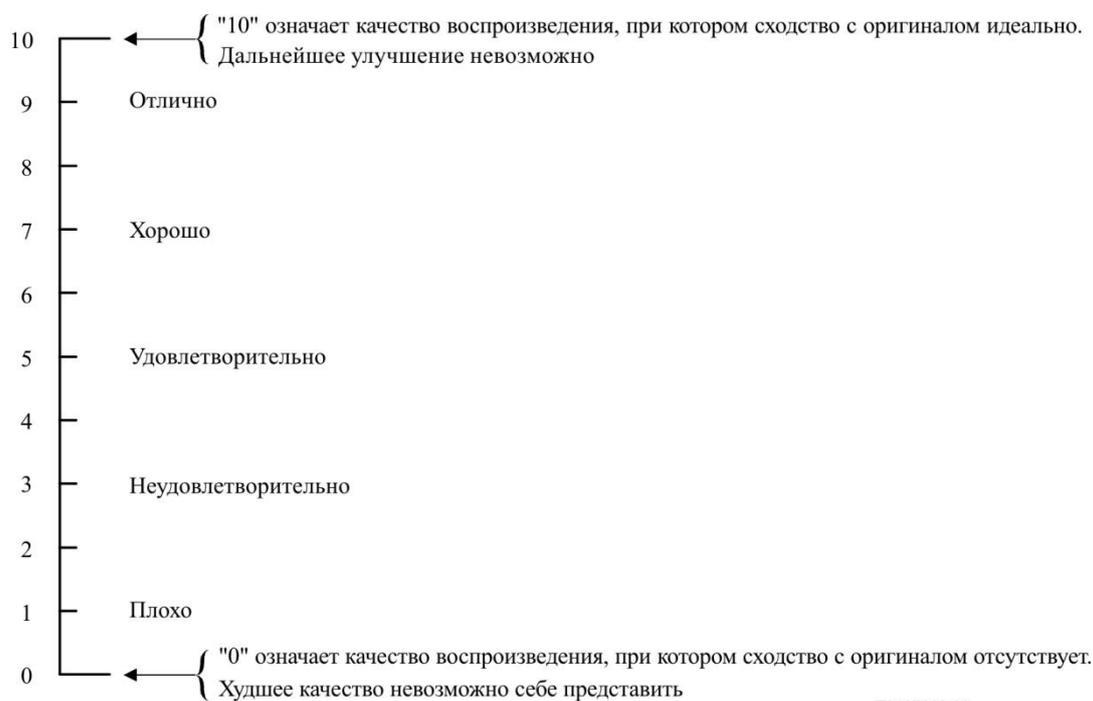


Рисунок В.2 – Одиннадцатibalльная числовая шкала качества

Для шкал обоих типов ответы наблюдателей могут фиксироваться либо в виде чисел, которые вписываются в анкету, либо как отметки на самой шкале (в этом случае в анкете должна быть приведена отдельная шкала для каждого оцениваемого условия). Когда требуются численные ответы, следует рекомендовать наблюдателям использовать десятичные дроби (например, 2,2 вместо 2), но у них может оставаться возможность использования только целых чисел.

Следует отметить, что могут возникнуть трудности с переводом наименования категорий шкалы на другие языки. При этом соотношения между категориями могут отличаться от их соотношения на языке оригинала [b-Virtanen].

Еще одна возможность – использование непрерывной шкалы.

Поскольку непрерывные данные обычно округляются до некоторой разумной точности, чтобы упростить сбор данных, можно использовать шкалу голосования, подобную той, которая показана на рисунке В.3. Ярлыки используются только в конечных точках, а в середине шкалы ставится отметка. Это должно уменьшить смещение, вызванное интерпретацией ярлыков. Каждая область может соответствовать конкретному численному значению, и данные собираются без неоднозначности.

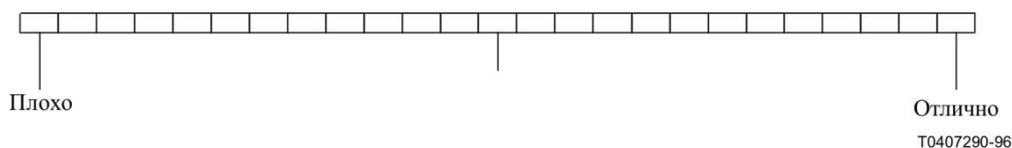


Рисунок В.3 Квазинепрерывная шкала оценки качества

В.2 Дополнительные размерности оценки

Если оцениваемые в ходе испытаний системы оказываются достаточно близкими по общему качеству и, следовательно, получают очень схожие оценки, может быть целесообразно оценить дополнительные компоненты качества по отдельной шкале для каждого условия. Таким образом можно получить информацию о конкретных характеристиках, по которым объекты испытаний воспринимаются как существенно различающиеся, несмотря на почти одинаковое общее качество. Результаты таких дополнительных тестов могут дать ценную диагностическую информацию о тестируемых системах.

Ниже приводятся примеры размерностей оценки, которые можно предложить для определения факторов, дополняющих общее воспринимаемое качество изображения, с указанием того, вносит ли данный фактор положительный или отрицательный вклад:

- яркость (положительный);
- контрастность (положительный);
- цветовоспроизведение (положительный);
- очертание контура (положительный);
- стабильность фона (положительный);
- скорость сборки изображения (положительный);
- подергивание (отрицательный);
- эффекты размытия (отрицательный);
- эффекты mosquito шума (отрицательный);
- двойные изображения/тени (отрицательный);
- гало (отрицательный).

Недавние исследования показали, что эти факторы можно объединить в прогнозируемое общее качество, придав каждому фактору соответствующие весовые коэффициенты, а затем сложив их вместе [b-RACE].

Чтобы оценить размерные показатели общего качества видеоизображения по отдельности, можно использовать специальную анкету. Примеры вопросов, которые могут задаваться после представления каждого тестового условия, приведены в нижеследующей анкете.

Анкета

Ответьте, пожалуйста, на следующие вопросы о последней показанной последовательности видеок кадров (видеоряде).

Вы можете высказать свое мнение, поставив отметку на нижеследующей шкале.

1) Как вы оцениваете цвета изображения?

2) Как вы оцениваете контрастность изображения?

3) Как вы оцениваете границы изображения?

4) Как бы вы оценили непрерывность движения?

5) Заметили ли вы какое-либо мерцание в видеоряде? Да Нет

Если вы заметили мерцание, оцените его, пожалуйста, по следующей шкале.

6) Заметили ли вы какое-либо размывание изображения в видеоряде? Да Нет

Если вы заметили размывание, оцените его, пожалуйста, по следующей шкале.

ПРИМЕЧАНИЕ. – При использовании этих шкал все учитываемые категории качества/ухудшений (например, непрерывность движения, мерцание, размывание и т. д.) следует обстоятельно пояснить на наглядном примере во время учебных занятий.

Приложение С

Одновременное представление пар последовательностей

(Данное Приложение является неотъемлемой частью настоящей Рекомендации)

С.1 Введение

Когда для оценки систем в ходе испытаний используется уменьшенный формат изображения, такой как CIF, QCIF, SIF и т. д., и метод DCR либо PC, может быть целесообразно демонстрировать последовательности каждой пары одновременно на одном и том же мониторе.

Преимущества одновременной демонстрации (SP):

- 1) SP значительно сокращает продолжительность испытаний;
- 2) при использовании изображения подходящих размеров наблюдателям легче оценить различия между источниками воздействия;
- 3) поскольку при одних и тех же условиях испытаний количество демонстраций сокращается вдвое, при использовании SP внимание наблюдателей обычно выше.

Одновременная демонстрация требует особых мер предосторожности во избежание предвзятости со стороны наблюдателей, вызванной типом демонстрации.

С.2 Синхронизация

Две последовательности должны быть полностью синхронизированы; это означает, что обе они должны запускаться и останавливаться на одном и том же кадре и что демонстрация должна быть синхронизирована. Это не исключает того, что могут сравниваться последовательности, кодированные с разными битовыми скоростями, при условии применения подходящей временной интерполяции.

С.3 Условия просмотра

Последовательности должны отображаться в двух окнах, расположенных бок о бок на фоне 50%-ого серого цвета (серый цвет определен в пункте 5.1), как показано на рисунке С.1. Чтобы уменьшить необходимость движения глаз при переключении внимания между двумя окнами, расстояние просмотра устанавливается равным $8H$, где H – высота изображения. Размер мониторов по диагонали должен составлять не менее 14 дюймов.

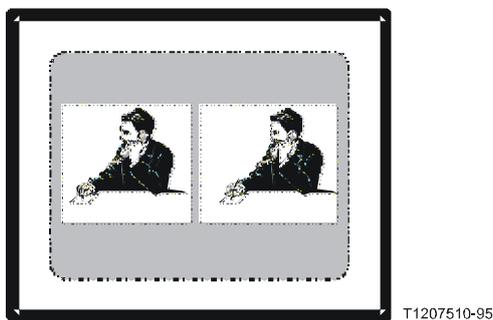


Рисунок С.1. Относительное положение двух последовательностей в SP

С.4 Демонстрация

В случае использования метода DCR эталон всегда должен размещаться с одной и той же стороны (например, слева), и наблюдатели должны знать относительное положение эталонных и тестовых условий.

В случае использования метода PC все пары последовательностей должны демонстрироваться в обоих возможных порядках (например, АВ, ВА). Это означает, что последовательности, которые были отображены с левой стороны, теперь отображаются справа, и наоборот.

Приложение D

Классы видеозображений и их атрибуты

(Данное Приложение является неотъемлемой частью настоящей Рекомендации)

В данной Рекомендации видеозображением самого высокого качества считается видеозображение [ITU-R BT.601], 8 бит/пиксель с линейным кодированием РСМ в формате 4 : 2 : 2, Y, C_R, C_B.

Таблица D.1. Определения классов видеозображения

TV 0	Без потерь: [ITU-R BT.601], 8 бит на пиксель, видеозображение, используемое для приложений без сжатия
TV 1	Используется для полного постпроизводства, много уровней редактирования и обработки, внутростанционная передача. Также используется для удаленного доступа к станции. Перцептивно прозрачно по сравнению с TV 0
TV 2	Используется для внесения простых изменений, небольшого количества редакций, наложения символов/логотипов, вставки программ и передачи между площадками. Примером трансляции может служить передача из сети в филиалы. Другими примерами являются нисходящий канал из региональной кабельной системы в местную головную станцию и высококачественная система видеоконференц-связи. Перцептивно почти прозрачно по сравнению с TV 0
TV 3	Используется для доставки изображений в дома/потребителям (без изменений). Другими примерами являются кабельная система трансляции из местной головной станции в дома и система видеоконференц-связи среднего или высокого качества. По сравнению с TV 2 присутствуют небольшие артефакты
MM 4	Все кадры кодированы. Небольшие артефакты по сравнению с TV 3. Видеоконференц-связь среднего качества. Обычно ≥ 30 кадров/с
MM 5	Кодер может пропускать кадры. Возможны ощутимые артефакты, но уровень качества пригоден для решения поставленных задач, например для низкокачественной видеоконференц-связи
MM 6	Ряд неподвижных изображений. Не предназначен для передачи движения (примеры: видеонаблюдение, графика)

Таблица D.2. Атрибуты классов видеозображения

Класс видеозображения	Пространственный формат	Частота кадров (Примечание 1)	Типичная вариация задержки (Примечание 2)	Номинальная битовая скорость видеозображения (Мбит/с)
TV 0	[ITU-R BT.601]	Макс. частота кадров	(Примечание 2)	270
TV 1	[ITU-R BT.601]	Макс. частота кадров	(Примечание 2)	18–50
TV 2	[ITU-R BT.601]	Макс. частота кадров	(Примечание 2)	10–25
TV 3	[ITU-R BT.601]	Макс. частота кадров со случайным повтором кадра	(Примечание 2)	1,5–8
MM 4a	[ITU-R BT.601]	~30 или ~25 кадров/с	Задержка < ≈150 мс Вариация < ≈50 мс	~1,5
MM 4b	CIF	~30 или ~25 кадров/с	Задержка < ≈150 мс Вариация < ≈50 мс	~0,7
MM 5a	CIF	10–30 кадров/с	Задержка < ≈1000 мс Вариация < ≈500 мс	~0,2
MM 5b	≤ CIF	1–15 кадров/с	Задержка < ≈1000 мс Вариация < ≈500 мс	~0,05
MM 6	CIF-16CIF	Предел → 0 кадров/с	Без ограничения	< 0,05, предел → 0 кадров/с

ПРИМЕЧАНИЕ 1. – Обычно 30 кадров/с для систем с 525 строками и 25 кадров/с для систем с 625 строками.
 ПРИМЕЧАНИЕ 2. – Все системы широковещательной передачи имеют постоянную, но не обязательно короткую одностороннюю задержку и постоянную вариацию задержки. В большинстве широковещательных приложений задержка будет короткой, например от 50 до 500 мс для высококачественной видеоконференц-связи, а в общем случае для приложений разговорного типа предпочтительна задержка менее 150 мс (см. [b-ITU-T G.114]). Допускается вариация задержки в пределах заданного диапазона, но она не должна приводить к нарушающим восприятие эффектам изменения масштаба времени.

Дополнение I

Тестовые последовательности

(Данное Дополнение не является неотъемлемой частью настоящей Рекомендации)

Ключевым моментом при планировании субъективной оценки является выбор подходящих тестовых последовательностей. Когда результаты испытаний, проводимых с различными группами наблюдателей или в разных лабораториях, должны коррелироваться между собой, важно, чтобы был доступен общий набор тестовых последовательностей.

Первый набор таких последовательностей описан в таблице I.1. В этой таблице для каждой последовательности указана следующая информация:

- категория (определенная в таблице A.1);
- краткое описание сцены;
- исходный формат (либо 625 или 525 строк, либо формат [ITU-R BT.601-4] или Betacam SP);
- значения пространственной и временной информации (определенной в пунктах 5.3.1 и 5.3.2 соответственно).

Все последовательности, перечисленные в таблице I.1, общедоступны и могут свободно использоваться для оценок и демонстраций. Некоторые из предложенных последовательностей относятся к библиотеке CCIR, описанной в [b-CCIR Report 1213].

Другие последовательности из библиотеки CCIR могут соответствующим образом использоваться для конкретных приложений, например приложений на основе хранения и извлечения видеозаписей.

Набор тестовых последовательностей продолжают изучать. Набор тестовых последовательностей, перечисленных в таблице I.1, можно улучшить или расширить как минимум двумя способами:

- 1) включить последовательности, отражающие более широкий диапазон приложений (мобильный видеофон, удаленная классная комната и т. д.);
- 2) сделать исходным форматом для каждой последовательности формат [ITU-R BT.601-4] в системах как с 525, так и 625 строками.

Таблица I.1. Тестовые последовательности для оценки качества видеоизображения в мультимедийных приложениях

Последовательность	Категория	Описание	Исходный формат	SI	TI
washdc	D	Карта Вашингтона (округ Колумбия) с движением рук и карандаша	Betacam SP (525 строк)	130,5	17,0
3inrow	C	Люди за столом, панорамирование камеры	Betacam SP (525 строк)	81,7	30,8
vtc1nw	A	Сидящая женщина читает новости	Betacam SP (525 строк)	56,2	5,3
Susie	A	Девушка говорит по телефону	ITU-R BT.601-4 525/625 строк	58,7	24,6
flower garden	E	Пейзаж, панорамирование камеры	ITU-R BT.601-4 525/625 строк	227,0	46,4
smity2	B	Продавец за прилавком с журналом	Betacam SP (525 строк)	154,5	35,1

Дополнение II

Инструкции по просмотру тестовых видеоизображений

(Данное Дополнение не является неотъемлемой частью настоящей Рекомендации)

Следующие разделы можно использовать в качестве основы инструкций для оценщиков, участвующих в экспериментах с применением методов ACR, ACR-HR, DCR или PC.

Кроме того, инструкции должны содержать информацию о приблизительной продолжительности испытаний, паузах, предварительных испытаниях и других деталях, полезных для оценщиков. Эта информация сюда не включена, поскольку она зависит от конкретной реализации.

II.1 ACR и ACR-HR

Доброе утро, и спасибо, что пришли.

В этом эксперименте вы увидите короткие видеоролики на экране, который находится перед вами. Вы должны оценить качество каждого из них, используя один из пяти уровней следующей шкалы.

- 5 Отлично
- 4 Хорошо
- 3 Удовлетворительно
- 2 Неудовлетворительно
- 1 Плохо

Прежде чем вынести суждение, внимательно просмотрите весь видеоряд.

II.2 DCR

Доброе утро, и спасибо, что пришли.

В этом эксперименте вы увидите короткие видеоролики на экране, который находится перед вами. Каждый из них демонстрируется два раза подряд: в каждой паре обрабатывается только второй видеоряд. После демонстрации каждой пары вы должны оценить степень ухудшения второго видеоряда по сравнению с первым. Выразите свое мнение, используя следующую шкалу.

- 5 Незаметное
- 4 Заметное, но не раздражающее
- 3 Немного раздражающее
- 2 Раздражающее
- 1 Сильно раздражающее

Прежде чем вынести суждение, внимательно просмотрите всю пару видеороликов.

II.3 PC

Доброе утро, и спасибо, что пришли.

В этом эксперименте вы увидите короткие видеоролики на экране, который находится перед вами. Каждый из них демонстрируется два раза подряд: каждый раз через другой кодек. Порядок демонстраций и комбинация кодеков в парах изменяются случайным образом. После демонстрации каждой пары вы должны выразить свое предпочтение, отметив одно из следующих полей. Отметьте

поле 1, если вы предпочитаете первый видеоряд, или поле 2, если вы предпочитаете второй видеоряд из пары.

1

2

Прежде чем вынести суждение, внимательно просмотрите всю пару видеороликов.

Дополнение III

Одновременное воздействие двух источников для непрерывной оценки

(Данное Дополнение не является неотъемлемой частью настоящей Рекомендации)

Метод одновременного воздействия двух источников для непрерывной оценки (SDSCE) подходит для оценки влияния на точность визуальной информации редких нарушений, таких как ошибки передачи. Этот метод основан на методе непрерывной оценки качества с одним источником воздействия (SSCQE), описанном в [ITU-R BT.500-9].

III.1 Процедура испытаний

Группа участников одновременно просматривает две последовательности – эталонную и соответствующую условиям испытаний. Если формат последовательностей SIF (формат стандартного изображения) или формат меньшего размера, то обе последовательности могут отображаться рядом на одном и том же мониторе, иначе должны использоваться два соседних монитора.

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

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

III.2 Этап обучения

Этап обучения – это важнейшая часть данного метода испытаний, поскольку участники могут неправильно понять свою задачу. Для уверенности в том, что все участники получают абсолютно одинаковую информацию, должны быть предоставлены письменные инструкции. Они должны включать пояснение о том, что участники увидят, что они должны оценить (то есть разницу в качестве) и как им следует выразить свое мнение. На все вопросы участников должны быть даны ответы, чтобы по мере возможности не допустить никакого расхождения в понимании с руководителем испытания.

После предоставления инструкций должен быть проведен демонстрационный сеанс. Это обеспечивает ознакомление участников с процедурами голосования и видами искажений.

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

После того как моделирование испытания завершено, экспериментатор должен убедиться, что когда условия испытаний эквивалентны эталонным условиям, оценки близки к ста; в противном случае экспериментатор должен повторить объяснение и моделирование испытания.

III.3 Особенности протокола испытаний

В описании протокола испытаний применяются следующие определения.

- *Фрагмент видеоизображения (video segment, VS)* – VS соответствует одной последовательности видеок кадров.
- *Условие испытания (test condition, TC)* – это может быть конкретный процесс, связанный с видеоизображением, условие передачи или то и другое вместе. Каждый VS должен обрабатываться в соответствии по крайней мере с одним TC. Кроме того, в список TC должны быть добавлены эталоны для образования оцениваемых пар эталон/эталон.

- *Сеанс (session, S)*. Сеанс – это последовательность различных пар VS/ТС без разделения, построенная в псевдослучайном порядке. В каждом сеансе хотя бы один раз встречаются все VS и ТС, но необязательно все сочетания VS/ТС. По всем сочетаниям VS/ТС должно голосовать одно и то же число наблюдателей (но необязательно одни и те же наблюдатели).
- *Испытательная демонстрация (test presentation, TP)*. Испытательная демонстрация – это последовательность сеансов, охватывающих все сочетания VS/ТС.
- *Период вынесения решений* – каждому наблюдателю предлагается непрерывно выносить решения в ходе сеанса.

III.4 Обработка данных

После проведения испытания получается один или несколько файлов, содержащих все решения, полученные в ходе разных сеансов (S), представляющих общее количество решений в отношении испытательной демонстрации (TP). Первая проверка достоверности данных может быть выполнена путем проверки того, что рассмотрена каждая пара VS/ТС и что по каждой из них вынесено равное количество решений.

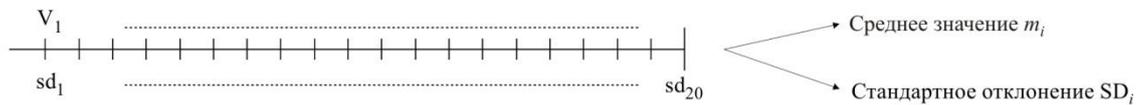
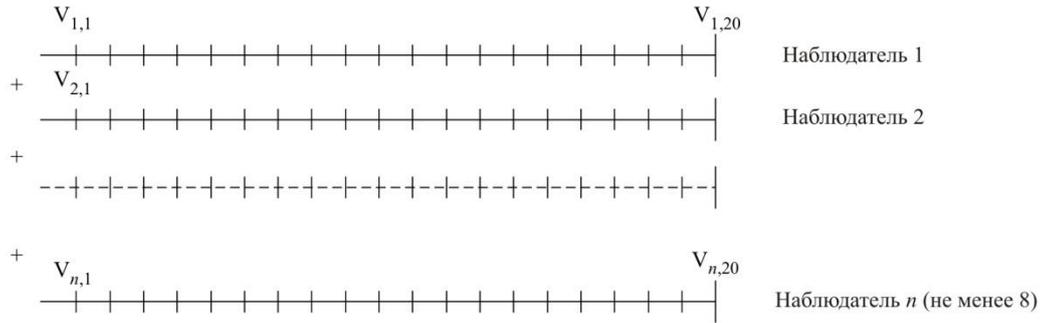
Данные, собранные в ходе испытаний, проведенных в соответствии с этим протоколом, могут быть обработаны тремя разными способами:

- статистический анализ каждого отдельного VS;
- статистический анализ каждого отдельного ТС;
- общий статистический анализ всех пар VS/ТС.

В каждом случае необходим многошаговый анализ.

- Средние значения и стандартные отклонения рассчитываются по каждому решению путем накопления наблюдателей, как показано на рисунке III.1.
- Затем каждый VS рассматривается как совокупность сегментов вынесения решений (голосования) максимальной продолжительностью 10 с. Поскольку на оценку последовательностей, которая длится не более 10 с, не влияет ни эффект новизны, ни эффект забывания, для каждого сегмента вынесения решения рассчитывается среднее и стандартное отклонение средних значений, вычисленных на предыдущем этапе, как показано на рисунке III.1. Когда требуется подробная информация о вариации качества, продолжительность сегмента вынесения решений должна быть короткой (около одной секунды). Результаты этого этапа могут быть представлены на временной диаграмме, как показано на рисунке III.2.
- Анализируется статистическое распределение средних значений, рассчитанных на предыдущем этапе (то есть соответствующих каждому сегменту вынесения решений), и частота их появления. Во избежание эффекта новизны, обусловленного предыдущими сочетаниями VS/ТС, первые десять секунд голосования для каждой выборки VS/ТС отбрасываются. Пример приведен на рисунке III.3.
- Общая характеристика раздражающего воздействия рассчитывается путем накопления данных о частоте появления. В этом расчете должны быть учтены доверительные интервалы, как показано на рисунке III.4. Общая характеристика раздражающего воздействия соответствует этой статистической интегральной функции распределения, указывая соотношение между средними значениями для каждого сегмента вынесения решений и соответствующей кумулятивной частотой появления.

- 1) Вычисление средней оценки (V) и стандартного отклонения (sd) на каждый случай голосования, по наблюдателям, на каждую последовательность для голосования для каждой комбинации VS/TS



- 2) Вычисление среднего значения (M) и стандартного отклонения (SD) на каждую последовательность для голосования продолжительностью 1 с для комбинации VS/TS

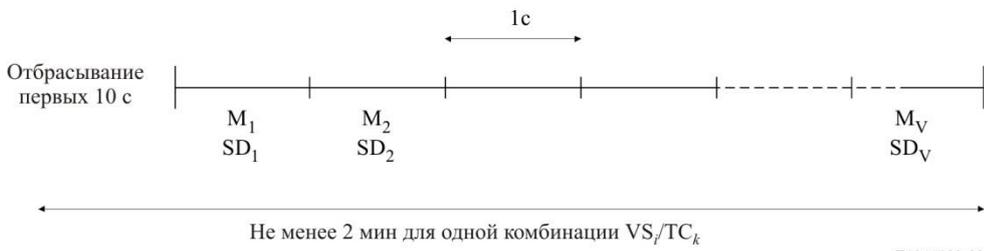


Рисунок III.1 – Обработка данных

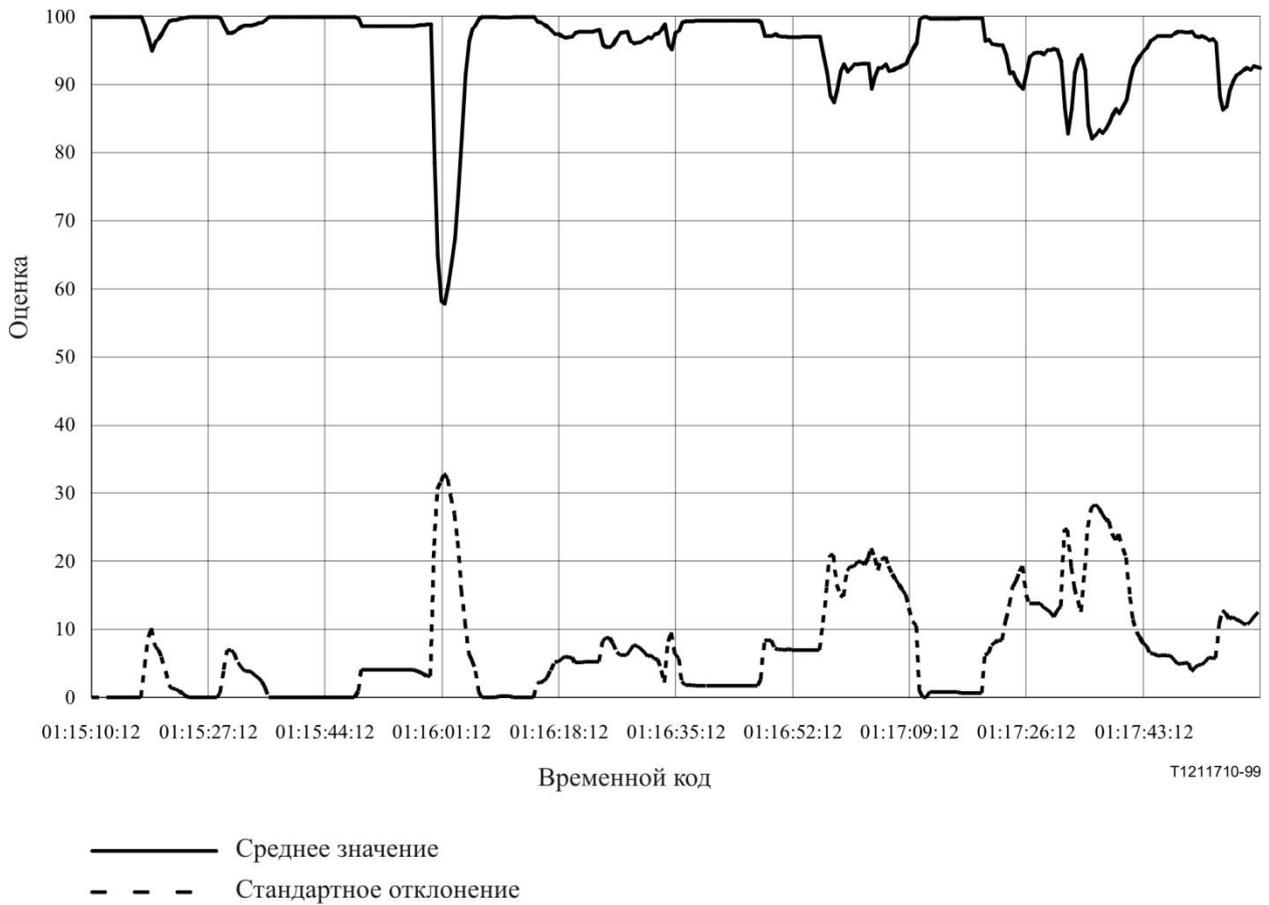


Рисунок III.2 – Исходная временная диаграмма

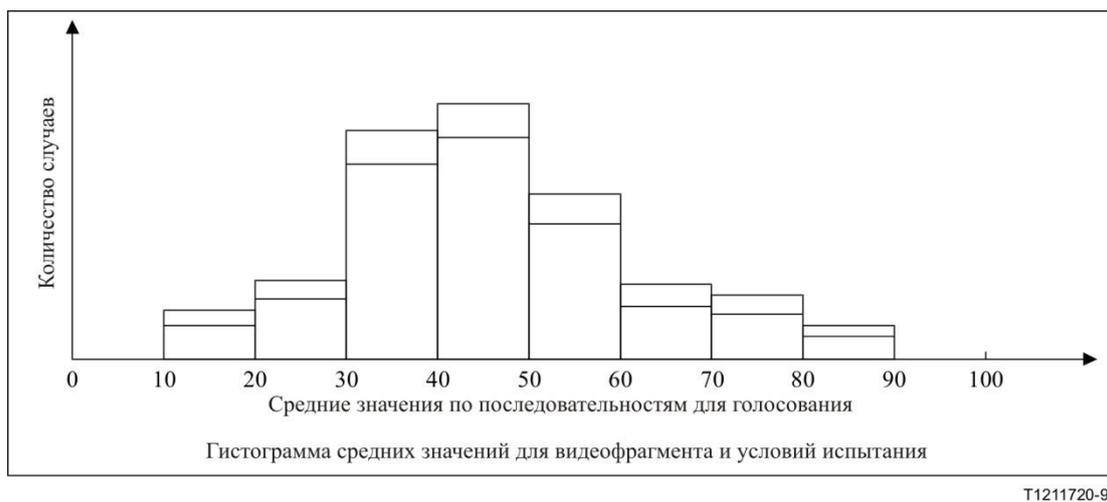


Рисунок III.3 – Связь между нарушениями и числом их проявлений

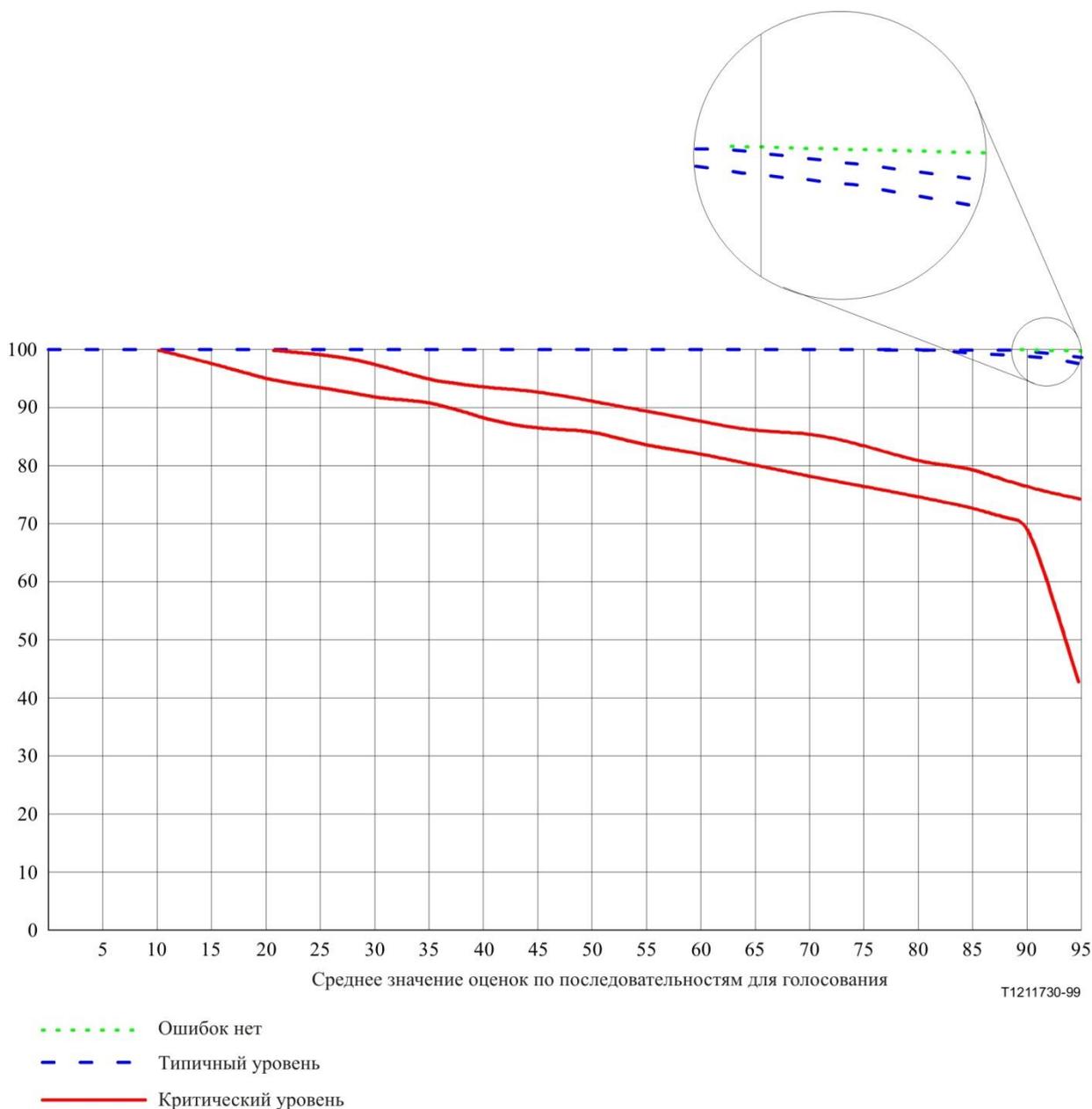


Рисунок 3.4 – Общие характеристики раздражающих факторов, рассчитанные на основе статистических распределений и с учетом доверительного интервала

3.5 Надежность участников

Надежность участников можно качественно оценить 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 тем, что описаны в [ITU-R BT.500-9] для метода SSCQE.

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

систематического 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;

местного 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.

Этих двух нежелательных эффектов (нетипичное поведение и отклонение от нормального порядка) можно избежать. Разумеется, весьма важным фактором является обучение участников. Однако должна быть возможность использования средства, позволяющего выявлять несоответствующих наблюдателей и при необходимости отстранять их.

Дополнение IV

Объектно ориентированная оценка

(Данное Дополнение не является неотъемлемой частью настоящей Рекомендации)

Объектно ориентированные функции должны оцениваться как для всей сцены, так и по отдельным объектам. Это связано с тем, что общая сцена, состоящая из независимо кодированных объектов, может использоваться так, как она была создана автором, но в некоторых случаях с ней также можно производить различные манипуляции, и каждый отдельный объект может использоваться совершенно в ином контексте. По этой причине важно обеспечить баланс между общим качеством всей сцены и качеством текстуры и контуров каждого отдельного объекта.

Поэтому объектно ориентированные функции (масштабируемость объекта и масштабируемость качества объектов) должны оцениваться в два этапа.

Оценка полного изображения. Это классический тест для всей последовательности, включающий все визуальные объекты (VO). В качестве метода оценки может быть выбран либо ACR (см. пункт 6.1), либо DCR (см. пункт 6.3) в зависимости от диапазона битовых скоростей и критичности исходных последовательностей.

Объектно ориентированная оценка (object-based evaluation, OBE). В этом тесте отображается только один VO на сером фоне, и наблюдателям предлагается оценить качество/ухудшение качества демонстрируемого VO (в соответствии с методом тестирования, используемым при оценке полного изображения). Необходимо указать процент битовой скорости, расходуемой на VO. Оцениваемый VO извлекается из той же кодированной последовательности, которая использовалась при оценке полного изображения.

На рисунке IV.1 показаны два теста, которые должны быть выполнены для оценки масштабируемости объекта.

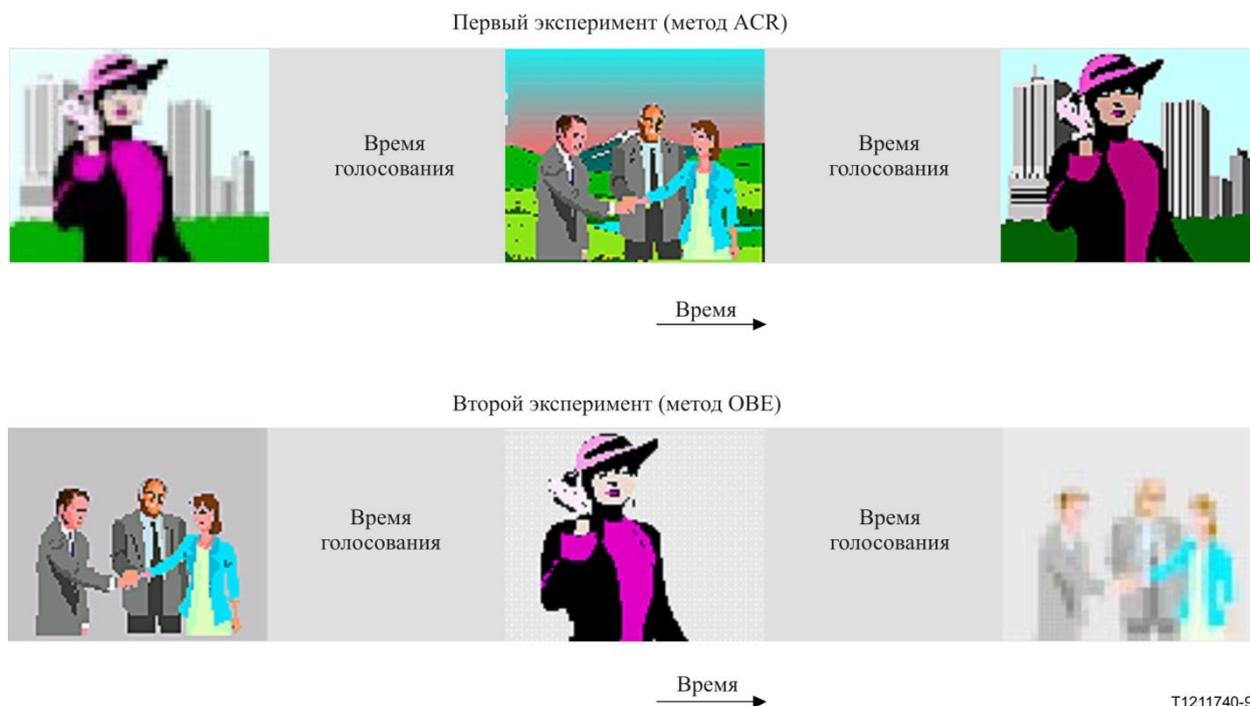


Рисунок IV.1 – Тесты для оценки масштабируемости объектов

В случае объектно ориентированной масштабируемости качества необходимо провести отдельные тесты для оценки пространственной масштабируемости и временной масштабируемости, и должен применяться только метод ОВЕ.

В случае как пространственной, так и временной масштабируемости ОВЕ должен применяться для оценки в одном и том же прогоне как VO, кодированных с "базовыми" битовыми скоростями, так и VO, кодированных с указанными улучшенными битовыми скоростями.

В целом оценка объектно ориентированных функций должна учитывать как качество всего кадра в целом, так и качество отдельных объектов. Первая оценка должна проводиться стандартными методами, вторая – методом ОВЕ.

Чтобы провести сравнение между разными системами на основе объектно ориентированного кодирования, экспериментатор должен заранее указать относительный вес, присвоенный общему качеству и качеству отдельных объектов.

В отдельных случаях также целесообразно вместо традиционных оценок качества использовать критерии оценки на основе задач. Например, при оценке системы удаленного мониторинга, которая будет использоваться в гараже, масштабируемость качества должна оцениваться с точки зрения разборчивости номерных знаков автомобилей. Задачи решаются экспериментатором в каждом конкретном случае в соответствии с целью эксперимента и видом исследуемого приложения.

Наконец, оценка качества объекта может применяться для исследования влияния качества отдельных объектов на общее качество сцены. Результаты такого исследования можно использовать для оптимизации схем объектно ориентированного кодирования.

Дополнение V

Дополнительная шкала оценок DCR

(Данное Дополнение не является неотъемлемой частью настоящей Рекомендации)

Можно использовать девятибалльную шкалу ухудшения качества, как показано на рисунке V.1. Оценка 8 по этой шкале соответствует порогу восприятия ухудшения, то есть уровню ухудшения, при котором наблюдатель не совсем уверен, что он ощущает ухудшение.

9	Незаметное
8	
7	Заметное, но не раздражающее
6	
5	Немного раздражающее
4	
3	Раздражающее
2	
1	Сильно раздражающее

Рисунок V.1 – Девятибалльная числовая шкала ухудшения качества

Библиография

- [b-ITU-T G.114] Рекомендация МСЭ-Т G.114 (2003 г.), *Время односторонней передачи*
- [b-ITU-T H.261] Recommendation ITU-T H.261 (1993), *Video codec for audiovisual services at $p \times 64$ kbit/s*
- [b-ITU-T P.920] Recommendation ITU-T P.920 (1996), *Interactive test methods for audiovisual communications*
- [b-ITU-T Handbook] ITU-T Handbook (1993), *Handbook on Telephony*, ITU, Geneva
- [b-ITU-R BT.812] Recommendation ITU-R BT.812 (1992), *Subjective assessment of the quality of alphanumeric and graphic pictures in Teletext and similar services*
- [b-ITU-R BT.815-1] Recommendation ITU-R BT.815-1 (1994), *Specification of a signal for measurement of the contrast ratio of displays*
- [b-CCIR Report 1213] CCIR Report 1213 (1990), *Test pictures and sequences for subjective assessments of digital codecs*, Annex to Volume XI, Part 1
- [b-Gonzalez] Gonzalez, R.C. and Wintz, P. (1987), *Digital Image Processing*, 2nd Edition, Addison-Wesley Publishing Co., Reading, Massachusetts
- [b-RACE] RACE Industrial Consortium Project 1018 HIVITS, WP B5, Picture Quality Measurement, 1988
- [b-Snellen] Snellen Eye Chart
- [b-Beck] *Pseudo Isochromatic Plates* (1940), engraved and printed by The Beck Engraving Co., Inc., Philadelphia and New York, United States
- [b-Kirk] Kirk, R.E. (1982), *Experimental Design – Procedures for the Behavioural Sciences*, 2nd Edition, Brooks/Cole Publishing Co., California
- [b-Virtanen] Virtanen, M.T., Gleiss, N. and Goldstein, M. (1995), *On the use of Evaluative Category Scales in Telecommunications*, Human Factors in Telecommunication Conference, Melbourne
- [b-Guilford] Guilford, P. (1954), *Psychometric methods*, McGraw-Hill, New York
- [b-ISO/IEC 11172] ISO/IEC 11172:1993, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*

СЕРИИ РЕКОМЕНДАЦИЙ МСЭ-Т

Серия А	Организация работы МСЭ-Т
Серия D	Общие принципы тарификации
Серия E	Общая эксплуатация сети, телефонная служба, функционирование служб и человеческие факторы
Серия F	Нетелефонные службы электросвязи
Серия G	Системы и среда передачи, цифровые системы и сети
Серия H	Аудиовизуальные и мультимедийные системы
Серия I	Цифровая сеть с интеграцией служб
Серия J	Кабельные сети и передача сигналов телевизионных и звуковых программ и других мультимедийных сигналов
Серия K	Защита от помех
Серия L	Конструкция, прокладка и защита кабелей и других элементов линейно-кабельных сооружений
Серия M	Управление электросвязью, включая СУЭ и техническое обслуживание сетей
Серия N	Техническое обслуживание: международные каналы передачи звуковых и телевизионных программ
Серия O	Требования к измерительной аппаратуре
Серия P	Качество телефонной передачи, телефонные установки, сети местных линий
Серия Q	Коммутация и сигнализация
Серия R	Телеграфная передача
Серия S	Оконечное оборудование для телеграфных служб
Серия T	Оконечное оборудование для телематических служб
Серия U	Телеграфная коммутация
Серия V	Передача данных по телефонной сети
Серия X	Сети передачи данных, взаимосвязь открытых систем и безопасность
Серия Y	Глобальная информационная инфраструктура, аспекты протокола Интернет и сети последующих поколений
Серия Z	Языки и общие аспекты программного обеспечения для систем электросвязи