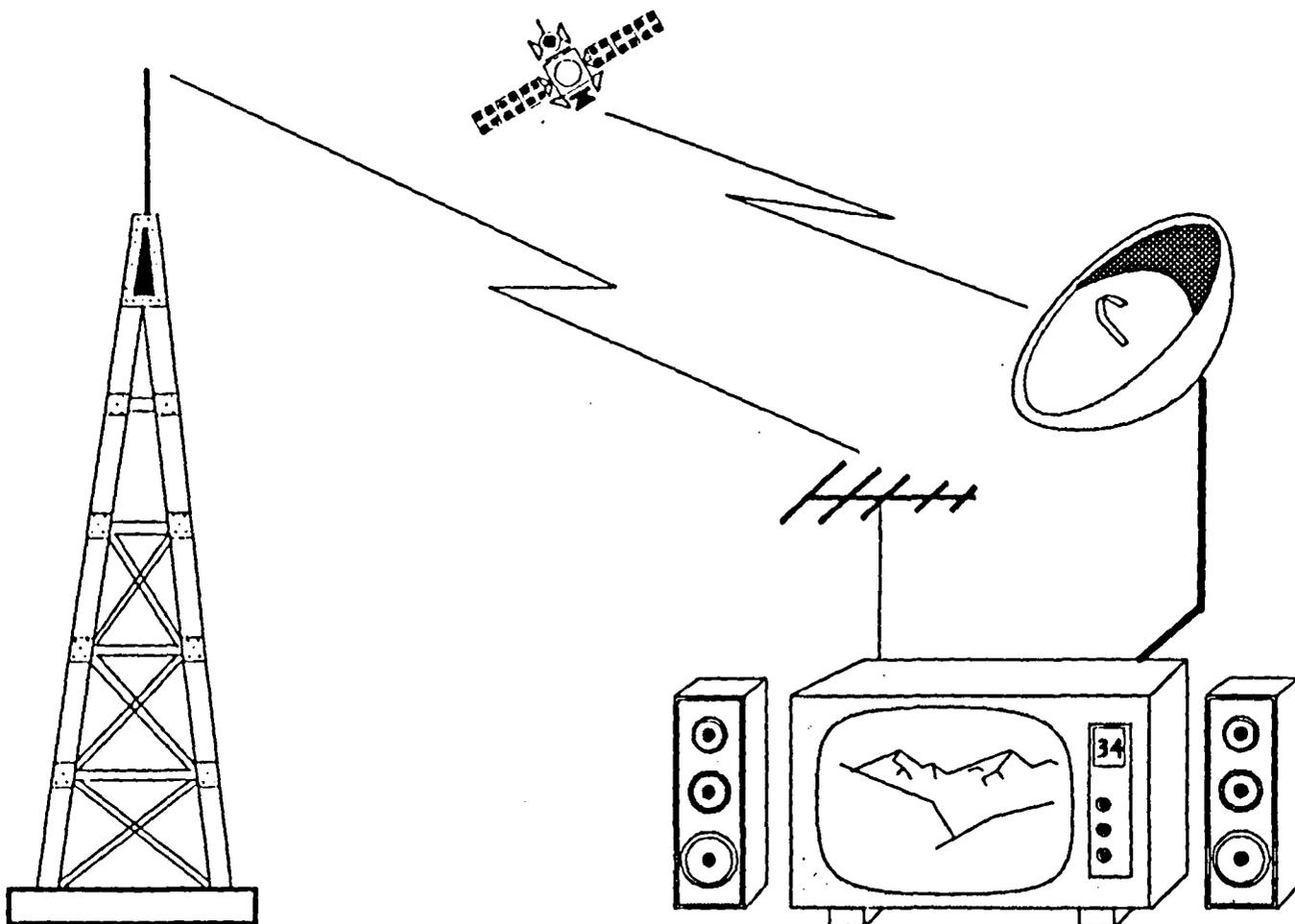




МЕЖДУНАРОДНЫЙ СОЮЗ ЭЛЕКТРОСВЯЗИ

# РЕКОМЕНДАЦИИ МККР, 1992 г.

(Новые и пересмотренные на 15 сентября 1992 г.)



Серия RBT

## ВЕЩАТЕЛЬНАЯ СЛУЖБА (ТЕЛЕВИДЕНИЕ)



МККР МЕЖДУНАРОДНЫЙ КОНСУЛЬТАТИВНЫЙ КОМИТЕТ ПО РАДИО

ISBN 92-61-04589-8

Женева, 1992 г.



© МСЭ 1992

Все права сохранены. Никакая часть данной публикации не может быть воспроизведена или использована в какой бы то ни было форме или с помощью каких-либо средств, электронных либо механических, включая изготовление фотокопий и микрофильмов, без письменного разрешения МСЭ.



## Recommendation 500-5 (1992)

### Method for subjective assessment of the quality of television pictures [Russian version]

Extract from the publication:  
*CCIR Recommendations: RBT series: Broadcasting Service (Television)*  
(Geneva: ITU, 1992), pp. 166-189

This electronic version (PDF) was scanned by the International Telecommunication Union (ITU) Library & Archives Service from an original paper document in the ITU Library & Archives collections.

La présente version électronique (PDF) a été numérisée par le Service de la bibliothèque et des archives de l'Union internationale des télécommunications (UIT) à partir d'un document papier original des collections de ce service.

Esta versión electrónica (PDF) ha sido escaneada por el Servicio de Biblioteca y Archivos de la Unión Internacional de Telecomunicaciones (UIT) a partir de un documento impreso original de las colecciones del Servicio de Biblioteca y Archivos de la UIT.

(ITU) للاتصالات الدولي الاتحاد في والمحفوظات المكتبة قسم أجراه الضوئي بالمسح تصوير نتاج (PDF) الإلكترونية النسخة هذه والمحفوظات المكتبة قسم في المتوفرة الوثائق ضمن أصلية ورقية وثيقة من نقلاً.

此电子版（PDF版本）由国际电信联盟（ITU）图书馆和档案室利用存于该处的纸质文件扫描提供。

Настоящий электронный вариант (PDF) был подготовлен в библиотечно-архивной службе Международного союза электросвязи путем сканирования исходного документа в бумажной форме из библиотечно-архивной службы МСЭ.

## РЕКОМЕНДАЦИЯ 500-5

МЕТОДИКА СУБЪЕКТИВНОЙ ОЦЕНКИ КАЧЕСТВА  
ТЕЛЕВИЗИОННЫХ ИЗОБРАЖЕНИЙ

(Вопрос 119/11)

(1974—1978—1982—1986—1990—1992)

МККР,

*учитывая,*

- a) что собран большой объем сведений о методиках, используемых в различных лабораториях для оценки качества изображений;
- b) что рассмотрение этих методик показывает, что различные лаборатории придерживаются в значительной мере общей точки зрения по ряду аспектов испытаний;
- c) что принятие стандартизованных методик имеет важнейшее значение при обмене информацией между различными лабораториями;
- d) что при регламентной или эксплуатационной оценке качества и/или ухудшения изображения с помощью пятибалльной шкалы качества и ухудшений, производимой во время повседневного или специального обслуживания инженерным персоналом, в определенной мере могут также применяться методики, рекомендованные для лабораторных испытаний;
- e) что введение новых видов обработки телевизионного сигнала, таких как цифровое кодирование и сокращение цифрового потока, новых видов телевизионных сигналов, использующих мультиплексированные во времени отдельные видеосигналы, и, возможно, новых служб, таких как телевидение повышенного качества и ТВЧ, могут потребовать изменения методик проведения субъективной оценки,

*рекомендует,*

1. чтобы для оценки качества изображений в ходе лабораторных экспериментов и во всех случаях, когда это возможно в условиях эксплуатации, использовались общие методики испытаний, шкалы оценок и условия наблюдений, описанные в последующих текстах;
2. чтобы в ближайшем будущем, несмотря на существование других методик и появление новых методик, по мере возможности использовались методики, описанные в § 2 и 3 приложения 1 к настоящей Рекомендации;
3. чтобы, учитывая важность создания основы для субъективной оценки, все отчеты по испытаниям были снабжены, по возможности, самыми полными описаниями схем испытаний, испытательных материалов, наблюдателей и методик.

*Примечание 1.* — Информация о методиках субъективной оценки для определения качественных показателей телевизионных систем содержится в приложении 1.

*Примечание 2.* — Информация о методиках субъективной оценки для определения искажений, обусловленных цифровым кодированием телевизионных сигналов, содержится в приложении 2.

## ПРИЛОЖЕНИЕ 1

## 1. Введение

Методики субъективной оценки используются для определения качественных показателей телевизионных систем с использованием измерений, которые гораздо лучше учитывают реакцию тех, кто мог бы быть наблюдателем систем, подвергаемых испытаниям. В связи с этим считают, что нельзя полностью охарактеризовать качественные показатели системы объективными средствами; следовательно, объективные измерения необходимо дополнить субъективными испытаниями.

Вообще, существуют два класса субъективных оценок. Во-первых, имеются оценки, определяющие качественные показатели систем в оптимальных условиях. Обычно их называют оценками качества. Во-вторых, имеются оценки, определяющие способность систем сохранять качество при неоптимальных условиях передачи или излучения. Обычно их называют оценками ухудшений.

Чтобы провести соответствующие субъективные испытания, на первом этапе необходимо из различных возможных вариантов выбрать такие, которые лучше всего подходят к целям и условиям конкретного вопроса оценки. На практике это требует принятия решения о выборе методики испытаний, испытательных материалов и условий наблюдения.

### 1.1 Выбор методики испытаний

При оценке в телевидении используется множество базовых методик испытаний. Однако на практике для решения конкретных вопросов оценки должны использоваться и конкретные методики. Перечень типичных задач оценки и методик, используемых для решения этих задач, приведен в таблице 1.

ТАБЛИЦА 1

Выбор методики испытаний

Задача оценки	Используемый метод	Источник (Рекомендация 500)
Измерение качества систем относительно эталона	Двухстимульный метод с непрерывной шкалой качества	§ 3
Квантификация качества систем (при отсутствии эталона)	Метод с использованием шкалы отношений <sup>(1)</sup>	§ 4
Сравнение качества различных систем (при отсутствии эталона)	Метод прямого сравнения или метод с использованием шкалы отношений <sup>(1)</sup>	§ 4
Нахождение факторов, определяющих различие систем по восприятию	Метод с использованием многомерной шкалы или метод факторного анализа	§ 4
Измерение различий между системами по конкретным факторам	Мультивалентный метод	
Измерение устойчивости систем (например, характеристик отказов)	Двухстимульный метод с использованием шкалы ухудшений	§ 2
Квантификация устойчивости систем (например, характеристик отказов)	Метод с использованием шкалы отношений <sup>(1)</sup>	§ 4
Определение момента, при котором ухудшение становится заметным	Оценка порога методом принудительного выбора или метод регулировки	§ 4
Определение различия систем по восприятию	Метод принудительного выбора	§ 4

<sup>(1)</sup> На основании некоторых исследований можно предположить, что этот метод более стабилен при оценке всего диапазона качества.

### 1.2 Выбор испытательных материалов

При оценке в телевидении для определения типа испытательных материалов существует несколько подходов. Однако на практике конкретные типы испытательных материалов должны использоваться для решения конкретных задач оценки. Перечень типичных задач оценки и испытательных материалов, используемых для решения этих задач, приведен в таблице 2.

### 1.3 Выбор условий наблюдения

При оценках в стандартном телевидении следует использовать один конкретный набор условий наблюдения. Однако для нормальной и критической оценок могут использоваться различные расстояния наблюдений (см. таблицу 3).

ТАБЛИЦА 2

## Выбор испытательного материала\*

Задача оценки	Используемый материал	Источник (Рекомендация 500)
Общие качественные показатели с использованием среднего материала	Общего характера, «критичный, но не чрезмерно»	§ 2
Возможности, критические применения (например, снабжение программами, постобработка и т. д.)	В широком диапазоне, включая весьма критичный материал для испытываемого применения	Приложение 1
Качественные показатели «адаптивных» систем	Материал, весьма критичный для используемого «адаптивного» метода	Приложение 1
Нахождение слабых мест и возможных улучшений	Критичный, ориентированный на отдельные признаки материал	
Нахождение факторов, определяющих различие систем по восприятию	Материал в широком диапазоне с очень богатым содержанием	
Взаимное преобразование различных стандартов	Критичный для различий (например, по частоте полей)	

Считают, что все испытательные материалы по замыслу могут быть частью содержания телевизионной программы. Дальнейшие руководящие принципы по выбору испытательных материалов содержатся в дополнениях 1 и 2.

ТАБЛИЦА 3

## Выбор условий наблюдений

Задачи оценки	Условия наблюдений	Источник (Рекомендация 500)
Оценка стандартных систем	Расстояние наблюдения равно 6 высотам изображения	§ 2
Оценка стандартных систем при критических условиях	Расстояние наблюдения равно 4 высотам изображения для 625-строчных систем и 4–5 высотам изображения для 525-строчных систем	§ 2

## 2. Двухстимульный метод с использованием шкалы ухудшений («метод ЕСВ»)

### 2.1 Общее описание

Типичные испытания могут потребовать либо оценки какой-то новой системы, либо определения влияния ухудшений в тракте передачи. Первоначальные шаги по организации испытаний должны включать выбор подходящего испытательного материала для обеспечения достоверности оценки и определение соответствующих условий испытания. Если интерес представляет влияние изменений параметров, то необходимо выбрать ряд значений этих параметров, которые охватывают весь диапазон возможных ухудшений небольшим числом приблизительно равных шагов. Если оценивается какая-то новая система, для которой значения параметров не могут изменяться, то необходимо либо вводить дополнительные ухудшения, но субъективно такие же, либо использовать другой метод, такой как в § 3.

Двухстимульный (ЕСВ) метод основан на использовании цикличности в том смысле, что сначала наблюдателю предъявляется неискаженный эталон, затем то же самое изображение с ухудшением качества. После этого наблюдателя просят оценить второе, помня первое. В сеансах, которые длятся до получаса, наблюдателю предъявляется несколько изображений или последовательностей изображений в случайном порядке и со случайно выбранными ухудшениями, охватывающими все требуемые комбинации. В оцениваемые изображения или последовательности включены и изображения без ухудшения качества. В конце серии сеансов вычисляется средняя оценка для каждого условия наблюдения и для каждого испытательного изображения.

В данном методе используется шкала ухудшений, при которой стабильность результатов оказывалась обычно выше для малых ухудшений, чем для больших. Хотя этот метод иногда использовался с ограниченными диапазонами ухудшений, более правильно использовать его при полном диапазоне ухудшений.

## 2.2 Общая схема

Обобщенная схема испытательной установки показана на рис. 1, ниже.

РИСУНОК 1

Обобщенная схема испытательной установки для двухстимульного метода с использованием шкалы ухудшений



Участники испытаний наблюдают видеомонитор для оценки, на который сигнал подается через хронированный переключатель. На вход этого переключателя сигнал может поступать либо непосредственно от источника сигнала либо через испытываемую систему. Наблюдателям предъявляют ряд испытательных изображений или последовательностей изображений. Они следуют попарно, причем первое в паре приходит непосредственно от источника, а второе является тем же самым изображением, прошедшим через испытываемую систему.

## 2.3 Сигналы источника

Сигнал источника создает эталонное изображение непосредственно и служит входным сигналом испытываемой системы. Такой сигнал должен обладать оптимальным качеством для используемого телевизионного стандарта. Отсутствие дефектов в эталонной части представленной пары — решающий фактор для получения устойчивых результатов.

Изображения и последовательности изображений, хранящиеся в памяти в цифровом виде, обеспечивают формирование сигналов источника, обладающих наибольшей воспроизводимостью, и именно поэтому они являются предпочтительными. Лаборатории могут обмениваться этими изображениями, для того чтобы сравнение систем было более значимым. Стандарт видеозаписи D-1 4:2:2 (Рекомендация 657) должен стать основой для обмена исходными изображениями и последовательностями, когда видеомагнитофоны этого стандарта будут широко распространены и станут экономически доступными. Возможно также использование стандартов точной записи для компьютерных накопителей.

В случае необходимости проведения испытаний в короткий срок теледиапроектор для 35-мм диапозитивов является предпочтительным источником неподвижных изображений. Его разрешающая способность достаточна для оценки стандартного телевидения. Колориметрия и другие характеристики киноплёнки могут стать причиной различного субъективного восприятия в сравнении с изображениями от студийной камеры. Если такие характеристики влияют на результаты, должны использоваться непосредственно студийные источники, хотя часто это связано с гораздо большими неудобствами. Как правило, теледиапроекторы должны подстраиваться от изображения к изображению для достижения как можно более высокого субъективного качества изображения, поскольку такая ситуация может иметь место на практике.

Оценку возможностей вторичной обработки часто производят с использованием квазилинейной цветовой рирпроекции. При работе в студии такая цветовая рирпроекция очень чувствительна к студийному освещению. Следовательно, при оценке предпочтительно пользоваться специальной парой диапозитивов с квазилинейной цветовой рирпроекцией, которая всегда будет давать высококачественные результаты. Если необходимо, в диапозитив переднего плана может быть введено движение.

## 2.4 Условия наблюдения

Условия наблюдения, создаваемые для участников испытаний, должны быть следующими:

### 2.4.1 Общие условия

a)	Отношение расстояния наблюдения к высоте изображения	4H и 6H*
b)	Пиковая яркость экрана	70 кд/м <sup>2</sup>
c)	Отношение яркости погашенного экрана (трубка заперта) к пиковой яркости	≤ 0,02
d)	Отношение яркости экрана при воспроизведении только уровня черного в совершенно темной комнате к пиковой яркости экрана	≈ 0,01
e)	Отношение яркости фона за видеомонитором к пиковой яркости изображения	≈ 0,15
f)	Постороннее освещение комнаты	слабое
g)	Цветность фона	D <sub>95</sub>
h)	Отношение телесного угла, охватывающего ту часть фона, которая отвечает настоящей спецификации, к телесному углу, охватывающему изображение	≥ 9

### 2.4.2 Особые условия

a)	Типичное число наблюдателей при расстоянии до каждого видеомонитора 4H	2 (для первой половины сеансов) 3 (для второй половины сеансов)
b)	Типичное число наблюдателей при расстоянии до каждого видеомонитора 6H	как указано выше
c)	Видеомонитор**	высококачественный, размер экрана 22—26 дюймов (50—60 см)
d)	Яркость и контраст экрана	установка по сигналу PLUGE (см. Рекомендацию 814)
e)	Типичное число наблюдателей для каждого видеомонитора	5 (2 при 4H и 3 при 6H для первого сеанса, 3 при 4H и 2 при 6H для следующего сеанса, и т. д.)
f)	Особенности помещения(ий), предназначенного(ых) для наблюдения	комната, 3 стены которой задрапированы белым материалом, а 4-я (задняя) стена задрапирована серым.

## 2.5 Испытательный сеанс

Сеанс должен продолжаться до получаса и включать приблизительно до 40 предъявлений (см. § 2.6).

Сеансы komponуются в группы по два, чтобы дать возможность всем наблюдателям видеть изображения или последовательности изображений при расстояниях 4H и 6H. Если для единственной пары сеансов оказывается слишком много условий наблюдения, то должны быть организованы последующие пары. Необходимо использовать случайный порядок предъявлений (например, на основе греко-латинских квадратов); однако порядок условий наблюдения должен быть таким, чтобы любые влияния усталости или адаптации при оценивании были сбалансированы от сеанса к сеансу. Некоторые предъявления могут повторяться от сеанса к сеансу для проверки совпадения оценок. Каждое условие наблюдения должно быть воспроизведено дважды в одном и том же сеансе.

\* Предпочтительное расстояние для оценки стандартных систем (625/50, 525/60) составляет 6H, однако размещение наблюдателей на расстоянии 4H также приемлемо при условии, что либо результаты будут даны отдельно, либо в полученных средних оценках не будет значительной разницы.

\*\* Если используется более одного помещения для наблюдения, видеомониторы должны быть хорошо согласованы.

Изображения и ухудшения должны предъявляться в псевдослучайной последовательности, причем предпочтительно в разном порядке для каждого сеанса. В любом случае одни и те же испытательные изображения или последовательности изображений никогда не должны предъявляться два раза подряд с одинаковыми или разными уровнями ухудшений.

Диапазон ухудшений следует выбирать таким образом, чтобы большинством наблюдателей использовались все оценки; желательно, чтобы общая средняя оценка (усредненная с учетом всех мнений, высказанных при эксперименте) оказалась близкой к трем баллам.

Сеанс должен продолжаться не более получаса, включая пояснения и подготовку; испытательный сеанс можно начать с показа нескольких изображений, демонстрирующих диапазон ухудшений; суждения об этих изображениях не должны учитываться в окончательных результатах.

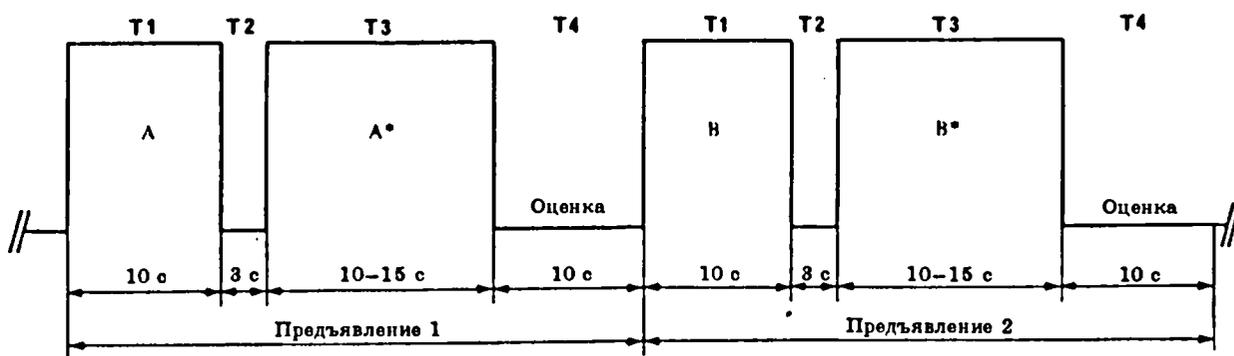
Дальнейшее рассмотрение вопроса о выборе уровней ухудшений приведено в дополнении 2.

## 2.6 Предъявление испытательного материала

Испытательный сеанс состоит из нескольких предъявлений. Последовательность предъявлений показана на рис. 2.

РИСУНОК 2

Последовательность предъявлений испытательного материала



А, В: эталонное изображение или последовательность  
А\*, В\*: испытательное изображение или последовательность

Каждое предъявление должно состоять из четырех фаз:

T1 = 10 с	эталонное изображение
T2 = 3 с	нейтральный серый фон, создаваемый видеосигналом с уровнем приблизительно 200 мВ
T3 = 10–15 с	испытательное условие
T4 = 10 с	нейтральный серый фон

Продолжительность фазы T3 может быть 10—15 с. Даже в случае движущихся изображений, как показывает опыт, продление этого периода более чем на 15 с не улучшает способности наблюдателей оценивать изображения.

## 2.7 Наблюдатели

Число наблюдателей должно быть по крайней мере 15 человек. Они не должны быть экспертами в том смысле, что не имеют дело непосредственно с качеством телевизионного изображения в их повседневной работе и не являются опытными наблюдателями. Перед сеансом с использованием специально отобранных таблиц наблюдатели должны быть проверены: они должны иметь нормальную или скорректированную до нормальной остроту зрения и нормальное цветовое зрение.

Предварительные результаты показывают, что наблюдатели-неэксперты могут представить более критичные оценки в случае предъявления изображений, создаваемых передающей и отображающей техникой повышенного качества.

## 2.8 Шкалы оценки

Следует использовать следующую пятибальную шкалу ухудшений:

- 5 незаметно
- 4 заметно, но не мешает
- 3 немного мешающее
- 2 мешает
- 1 очень мешает

Наблюдатели должны использовать бланк с ясно показанной шкалой, а также иметь пронумерованные ячейки таблицы или какие-то другие способы записи оценок.

## 2.9 Выбор испытательного материала

Некоторые параметры могут вызывать одинаковую степень ухудшений для большинства изображений или последовательностей изображений. В таких случаях результаты, полученные при очень малом числе изображений или последовательностей (например, при двух), все же могут дать довольно достоверную оценку.

Однако в новых системах часто возникают ухудшения, которые во многом зависят от содержания сюжета или последовательности. В таких случаях существует статистическое распределение вероятностей ухудшений и содержания изображений или последовательностей для всего множества программных часов. Без информации о форме такого распределения, что обычно имеет место, выбор испытательного материала и интерпретация результатов должны производиться очень осторожно.

Как правило, важно включать критичный материал, поскольку это можно учесть при интерпретации результатов, однако экстраполяция из некритичного материала невозможна. В случаях, когда содержание сюжета или последовательности влияет на результаты, материал, выбираемый для испытания системы, должен быть «критичным, но не чрезмерно». Выражение «не чрезмерно» говорит о том, что изображения могут быть, по замыслу, частью нормальной программы. В таких случаях необходимо использовать по крайней мере четыре категории материала: например, материал, половина которого явно критична, а другая половина — умеренно критична.

Разработкой испытательных неподвижных изображений и последовательностей изображений занимается ряд организаций. Есть надежда, что в будущем эта работа будет организована в рамках МККР.

МККР предложил материал для оценки цифровых систем, которые осуществляют сокращение цифрового потока сигналов, согласно Рекомендации 601, до значений 30—33 Мбит/с. Для оценки этих систем требуется обеспечить возможность выполнения различных операций вторичной обработки, таких как квазилинейная цветовая рирпроекция. В подобных случаях аппаратуру цветовой рирпроекции необходимо включить как в тракт «прямого» сигнала, так и в тракт сигнала испытательной системы. Эти сигналы затем могут быть предъявлены для оценки. Однако при использовании данного метода следует избегать включения тех эталонных изображений или последовательностей, качество которых уже понижено. Если интерес представляет оценка дополнительного понижения качества уже ухудшенного изображения, то они оба должны использоваться как испытательные последовательности.

Дальнейшие соображения по выбору испытательных материалов приведены в дополнениях 1 и 2.

## 2.10 Подготовка к испытаниям

Наблюдатели должны быть хорошо ознакомлены с методикой испытаний и типами ухудшений, которые могут встретиться при испытаниях. Возможны ответы на вопросы для достижения более четкого понимания, однако инструкции не должны изменяться от сеанса к сеансу, и во избежание навязывания мнения необходимо осторожно отвечать на вопросы.

В начале каждого сеанса необходимо ознакомить наблюдателей с типом испытания, шкалой оценки, последовательностью и временными рамками (эталонное изображение, уровень серого, испытательное изображение, моменты выставления оценки). Степень и тип ухудшений, которые подлежат оценке, необходимо пояснить изображениями, отличными от используемых в испытании, но со сравнимой чувствительностью к искажениям. Не должно подразумеваться, что самое низкое наблюдаемое качество обязательно соответствует самой низкой субъективной оценке. Необходимо просить наблюдателей основывать их суждения на общем впечатлении от данного изображения, и выражать эти суждения с использованием формулировок, принятых для описания субъективной шкалы.

Наблюдателей следует просить рассматривать изображения в течение полной продолжительности фаз Т1 и Т3. Выставление оценок должно производиться только в течение фазы Т4.

## 2.11 Представление результатов

Коррелированность результатов следует проверить путем исследования оценок, данных одним и тем же наблюдателем одному и тому же изображению в одном и том же сеансе. Если эти оценки различаются на два или более баллов, оба результата должны быть исключены.

В случае каждого испытываемого параметра должны быть определены среднее значение и среднеквадратическое отклонение для статистического распределения оценок испытания. Если испытания заключаются в определении зависимости искажения от изменяющегося параметра, то следует использовать методы аппроксимации кривых. Аппроксимированная логистическая кривая и логарифмическая шкала координат позволяют изобразить результаты в виде прямой линии, что является предпочтительной формой представления.

Результаты должны быть представлены вместе со следующими сведениями:

- подробности схемы испытания,
- детальная информация об испытательном материале,
- тип источника изображения и видеомониторов,
- число и тип наблюдателей,
- используемые эталонные системы,
- общая средняя оценка в эксперименте,
- начальные и скорректированные средние оценки и среднеквадратические отклонения, если один или несколько наблюдателей были исключены в соответствии с приведенной ниже процедурой.

После сеанса испытаний должны быть рассчитаны средние значения  $E(X_j)$  и среднеквадратические отклонения  $\sigma(X_j)$  для каждого уровня ухудшения или для каждой системы обработки, соответствующие оценке ( $j$ ). Эти средние значения основаны на распределении, двумя переменными которого являются сюжеты и наблюдатели. Затем должно быть установлено, является ли это распределение нормальным или нет с помощью теста  $\beta_2$  (расчет куртозисного коэффициента функции, то есть отношения момента четвертого порядка к квадрату момента второго порядка). Если значение  $\beta_2$  находится между 2 и 4, то распределение может быть принято за нормальное. Оценки  $X_{ij}$  каждого распределения  $j$  необходимо затем сравнить с соответствующим средним значением, увеличенным на соответствующее среднеквадратическое отклонение, умноженное на два (если распределение нормальное) или на  $\sqrt{20}$  (если распределение не является нормальным),  $P_i$ , и сравнить с соответствующим средним значением, уменьшенным на соответствующее среднеквадратическое отклонение, умноженное на два или  $\sqrt{20}$ ,  $Q_i$ . Каждый раз, когда оценка наблюдателя находится выше или ниже этих пределов, это должно быть зарегистрировано в документе каждого соответствующего наблюдателя; для оценок, превышающих указанные пределы значений выше ( $P_i$ ) и ниже ( $Q_i$ ), должны использоваться два различных расчета. Наконец, должны быть рассчитаны два следующих отношения:  $P_i + Q_i$  с учетом всех оценок каждого наблюдателя за весь сеанс, и  $P_i - Q_i$  к  $P_i + Q_i$  как абсолютная величина. Если первое больше 5%, а последнее меньше 30%, то наблюдатель  $i$  должен быть исключен из расчетов\*.

Описанная выше процедура может быть также выражена математически:

$$\text{если } X_{ij} \geq E(X_j) + 2 \cdot \sigma(X_j) \quad (\text{нормальное распределение})$$

или

$$\text{если } X_{ij} \geq E(X_j) + \sqrt{20} \cdot \sigma(X_j) \quad (\text{ненормальное распределение}),$$

то

$$P_i = P_i + 1.$$

$$\text{Если } X_{ij} \leq E(X_j) - 2 \cdot \sigma(X_j) \quad (\text{нормальное распределение})$$

или

$$\text{если } X_{ij} \leq E(X_j) - \sqrt{20} \cdot \sigma(X_j) \quad (\text{ненормальное распределение}),$$

то

$$Q_i = Q_i + 1.$$

$$\text{Если } \frac{P_i + Q_i}{\text{общее число оценок одного наблюдателя}} > 0,05 \text{ и } \left| \frac{P_i - Q_i}{P_i + Q_i} \right| < 0,3,$$

то наблюдатель  $i$  исключается из расчетов.

\* Эта процедура не должна применяться более одного раза к результату данного эксперимента. Более того, использование процедуры должно ограничиваться случаями, в которых принимает участие относительно немного наблюдателей (например, менее 20), причем все они не должны быть экспертами.

### 3. Двухстимульная методика с использованием непрерывной шкалы качества

#### 3.1 Общее описание

Типичное испытание заключается в оценке новой системы или влияния канала передачи на качество. Двухстимульная методика считается особенно полезной, когда невозможно создать для стимулов такие испытательные условия, которые охватывали бы весь диапазон качества.

Метод основан на использовании цикличности в том смысле, что наблюдателю предъявляют для просмотра пару изображений, полученных от одного и того же источника, однако одно из них подвергается процессу, подлежащему испытанию, а другое поступает непосредственно от источника. При этом наблюдателя просят оценить качество обоих изображений.

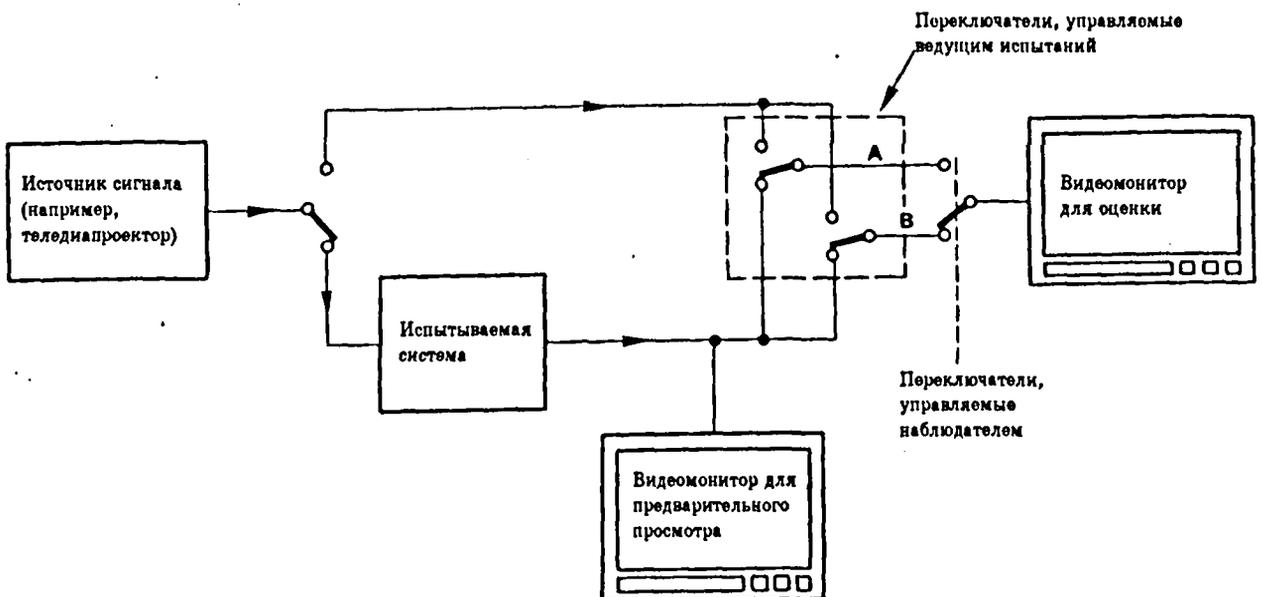
В сеансе, продолжительностью до получаса, наблюдателю предъявляют в случайном порядке пары изображений (также произвольные) и случайные ухудшения, охватывающие все необходимые сочетания. В конце сеанса рассчитываются средние оценки для каждого условия наблюдения и для каждого испытательного изображения.

#### 3.2 Общая схема

Обобщенная схема испытательной установки показана на рис. 3, ниже.

РИСУНОК 3

Обобщенная схема испытательной установки для двухстимульной методики с непрерывной шкалой качества



Существует два варианта этого метода, (I) и (II), описываемые ниже.

- (I) Наблюдатель, обычно одиночный, имеет возможность многократно переключать условия A и B до тех пор, пока он не составит мнение о каждом из них. На выход линий A и B эталонное изображение поступает либо непосредственно, либо через испытываемую систему, однако выбор подаваемого изображения и линии изменяются случайным образом при изменении условий, что фиксируется ведущим испытаний, но не объявляется.
- (II) Наблюдателям последовательно предъявляются изображения с выхода линий A и B, чтобы они составили свое мнение о каждом из них. Изображения на выходе линий A и B поступают для каждого предъявления так же, как в варианте (I), описанном выше. Стабильность результатов этого варианта при ограниченном диапазоне качества является предметом дальнейших исследований.

#### 3.3 Сигналы источника

Как и при использовании методики, описанной в § 2, однако искаженный эталон может не иметь такого же влияния на стабильность.

### 3.4 *Условия наблюдения*

Как и при использовании методики, описанной в § 2. Однако при использовании варианта (I) число наблюдателей, приходящихся на один видеомонитор, равно единице.

### 3.5 *Испытательный сеанс*

Как и при использовании методики, описанной в § 2. Нет необходимости по крайней мере в случае варианта (I) добиваться общей средней оценки 3.

### 3.6 *Предъявление испытательного материала*

Испытательный сеанс состоит из нескольких предъявлений. В случае варианта (I), в котором участвует один наблюдатель, для каждого предъявления наблюдатель имеет возможность переключать сигналы А и В до тех пор, пока он мысленно не даст оценку качеству, соответствующему каждому сигналу. Наблюдатель может, как правило, делать это два или три раза за период времени, не превышающий 10 с. В случае варианта (II), в котором одновременно участвуют несколько наблюдателей, перед регистрацией результатов каждая пара состояний предъявляется один или более раз в течение равных промежутков времени, с тем чтобы каждый наблюдатель мог мысленно оценить его качество, соответствующее этим состояниям, затем каждая пара состояний предъявляется повторно один или несколько раз с регистрацией результатов. Число повторений зависит от продолжительности испытательных последовательностей. В случае неподвижных изображений, вероятно, достаточны последовательности продолжительностью 3-4 с и пять повторений (с регистрацией оценок во время последних двух). В случае движущихся изображений с изменяющимися во времени дефектами, вероятно, достаточно использовать последовательности продолжительностью 10 с с двумя повторениями (с регистрацией оценок во время второй).

Если по практическим соображениям продолжительность имеющихся в наличии последовательностей составляет менее 10 с, то могут использоваться составные последовательности, включающие более короткие последовательности в качестве сегментов, что позволяет увеличить время предъявления до 10 с. Чтобы минимизировать разрывы на стыках, следующие друг за другом сегменты последовательности могут быть реверсированы во времени (иногда это называют «полидромным» отображением). Необходимо следить за тем, чтобы гарантировать, что условия испытаний, создаваемые при отображении реверсированных сегментов, были причинными процессами, то есть они должны создаваться посредством прохождения реверсированного во времени сигнала источника через испытываемую систему.

### 3.7 *Наблюдатели*

Как и при использовании методики, описанной в § 2.

### 3.8 *Шкала оценок*

В рассматриваемой методике требуется оценить два варианта каждого испытательного изображения. В каждой паре испытательных изображений одно без ухудшения, а другое может иметь, а может и не иметь ухудшения. Неискаженное изображение включается в качестве эталона, однако наблюдателю не сообщают, какое изображение является эталонным. В процессе испытаний положение эталонного изображения изменяется псевдослучайным образом.

Наблюдателей просто просят оценить общее качество каждого предъявленного изображения, поставив отметку на вертикальной шкале. Вертикальные шкалы напечатаны парами в соответствии с парным предъявлением каждого испытательного изображения. Такие шкалы обеспечивают систему непрерывной оценки, что позволяет избежать ошибок «квантования», тем не менее они разделены на пять равных по длине отрезков, соответствующих нормальной пятибалльной шкале качества МККР. Соответствующие словесные оценки для определенных различных уровней качества такие же, как обычно; однако в данном случае они включаются только для общего руководства и напечатаны лишь слева от первой шкалы в каждом ряду десяти двойных столбцов на бланке оценок. Рис. 4 показывает фрагмент типичного бланка оценок. Любая возможность путаницы между делениями шкалы и результатами испытания исключена, поскольку шкалы напечатаны синим, а записываемые результаты — черным.

### 3.9 *Выбор испытательного материала*

Как и при использовании методики, описанной в § 2.

### 3.10 *Подготовка к испытанию*

Как и при использовании методики, описанной в § 2, за исключением последнего абзаца в § 2.10.

РИСУНОК 4

Фрагмент бланка для оценки качества с использованием непрерывных шкал

	27		28		29		30		31	
	A	B	A	B	A	B	A	B	A	B
Отличное										
Хорошее										
Удовлетворительное										
Плохое										
Очень плохое										

### 3.11 Представление результатов

Возможны два различных подхода:

- Согласно первому результаты могут быть выражены в форме сравнительного испытания, то есть прямо указать изменение качества относительно эталонного изображения. Для каждого параметра испытаний должны быть даны среднее значение и среднеквадратическое отклонение статистического распределения измеренной разницы.
- Согласно второму (предпочтительному методу представления) результаты могут быть преобразованы в меры, используемые для описания эквивалентной оценки качества. Пары результатов (эталонных и испытательных) для каждого отдельного испытательного условия преобразуются посредством измерения соответствующей длины на бланке оценок в нормализованные оценки в пределах от 0 до 100. Для каждой испытываемой системы полученные оценки затем усредняются для различных групп наблюдателей, различных расстояний наблюдения и различных испытательных изображений; в итоге получают средние оценки для эталонных и испытательных условий для каждого сочетания переменных.

Поскольку средние оценки в случае эталонных условий всегда меньше единицы, необходима операция пересчета оценок испытаний. Пересчет производится вычитанием остаточного ухудшения. Средняя оценка для эталонного условия трактуется как остаточное ухудшение. Результаты вычитания выражаются в единицах ухудшения ( $imr$ ), однако при необходимости они могут быть снова преобразованы в средние оценки.

Отчет должен включать такую же дополнительную информацию, как и при использовании методики, описанной в § 2, за исключением средней оценки.

## 4. Другие методики оценки

При соответствующих обстоятельствах должна применяться одностимульная методика или методика сравнения стимулов.

### 4.1 Одностимульные методики

При использовании одностимульных методик предъявляется одиночное изображение или последовательность изображений, а наблюдатель проставляет показатель качества для всего предъявления.

#### 4.1.1 Наблюдатели

Для испытаний, проводимых в лабораторных условиях, наблюдатели обычно выбираются так, как описано в § 2.7. Необходимо, чтобы число наблюдателей зависело от чувствительности и надежности принятой процедуры испытаний и от ожидаемой степени ухудшения при исследуемом условии. При нормальных обстоятельствах в одном испытании участвует 10—20 наблюдателей.

#### 4.1.2 Испытательные изображения

Для испытаний, проводимых в лабораторных условиях, содержание испытательных изображений должно выбираться так, как описано в § 2.9.

После выбора содержания испытательных изображений их подготавливают таким образом, чтобы они отражали различные варианты рассматриваемых систем или диапазон(ы) изменения одного (или нескольких) факторов. Если рассматриваются два или более факторов, то изображения могут быть подготовлены двумя способами. Согласно первому каждое изображение представляет один уровень только одного фактора. Согласно второму каждое изображение представляет один уровень каждого из рассматриваемых факторов, однако с учетом всех изображений каждый уровень каждого фактора встречается вместе с каждым из уровней всех других факторов. Оба метода позволяют четко связать результаты с каждым из конкретных факторов. Второй метод позволяет дополнительно обнаружить взаимосвязь факторов (например, неаддитивные эффекты).

#### 4.1.3 Условия наблюдения

Как было замечено, когда наблюдатели остаются одни со своими собственными устройствами, они могут выбирать большие расстояния наблюдения, чем используемые при субъективных испытаниях. Соотношение между предпочтительными расстояниями наблюдения и теми, которые используются в испытаниях, требует дальнейшего изучения.

#### 4.1.4 Испытательный сеанс

До начала испытательного сеанса наблюдателям описываются задача наблюдения и, как правило, примеры изображений или последовательностей изображений. Инструкции обычно даются в письменной форме или в записи. Соблюдается осторожность, чтобы избежать оказания влияния на наблюдателей при выполнении их задачи.

Сеанс состоит из серии оценочных испытаний. Последние должны производиться в случайном порядке и предпочтительно в различном случайном порядке для каждого наблюдателя. При использовании единой случайной последовательности ведущий обычно следит за тем, чтобы одно и то же изображение не предъявлялось дважды подряд с одинаковым видом и уровнем ухудшения.

Типичное оценочное испытание предусматривает предъявление трех изображений: неяркого серого поля адаптации, поля стимула и неяркого серого поля постпредъявления. Продолжительность показа этих изображений меняется в зависимости от задачи наблюдения, материалов (например, неподвижное или движущееся изображение) и рассматриваемых вариантов или факторов, однако продолжительности 3, 10 и 10 с соответственно не являются необычными. Показатель (показатели) качества наблюдатель может проставить во время отображения либо поля стимула, либо поля постпредъявления.

#### 4.1.5 Виды одностимульных методик

В телевизионных субъективных испытаниях в основном используются три вида одностимульных методик.

##### 4.1.5.1 Методики оценки категорий

При вынесении суждений о категориях наблюдатели относят изображение или последовательность изображений к одной из набора категорий, которые, как правило, определяются семантически. Категории могут отражать суждение о том, обнаружен или нет какой-либо атрибут (например, для установления порога ухудшения). Шкалы оценки категорий, предназначенные для оценки качества изображений или ухудшения изображений, использовались наиболее часто; такие шкалы МККР приведены в таблице 4, ниже. В процессе эксплуатационного контроля иногда используются половинные градации оценок. В специальных случаях используются шкалы, предназначенные для оценки удобочитаемости текста, напряжения при чтении и полезности изображений.

Данный метод дает распределение суждений по шкале категорий для каждого испытательного условия. Способ анализа получаемых ответов зависит от собственно суждения (обнаружение и т.д.) и желаемой информации (порог обнаружения, ранг или центральная тенденция испытательных условий, психологическое «расстояние» между различными условиями). Известны многие методы анализа.

ТАБЛИЦА 4

## Шкалы качества и ухудшений МККР

Пятибалльная шкала	
Качество	Ухудшение
5 Отличное	6 Незаметно
4 Хорошее	4 Заметно, но не мешает
3 Удовлетворительное	3 Немного мешает
2 Плохое	2 Мешает
1 Очень плохое	1 Очень мешает

**4.1.5.2 Методики без оценки категорий**

При вынесении суждений без оценки категорий наблюдатели определяют значение для каждого показанного изображения или каждой последовательности изображений. Имеются две разновидности этой методики.

Согласно методике на основе непрерывной шкалы, которая является вариантом методики оценки категорий, наблюдатель определяет для каждого изображения или последовательности изображений отметку на линии, проведенной между двумя словесными отметками (например, между концами шкалы категорий, представленной в таблице 4). Такая шкала может иметь дополнительные словесные отметки в промежуточных точках для ссылки. Расстояние от конца шкалы принимается как показатель качества для каждого условия испытаний.

Согласно методике на основе числовой шкалы наблюдатель определяет для каждого изображения или последовательности изображений некое число, которое отражает оцененный уровень специфицированного параметра (например, резкости изображения). Диапазон этих чисел может быть как ограниченным (например, 0—100), так и неограниченным. Иногда назначенное число отражает суждение об уровне в «абсолютных» понятиях (без прямой ссылки на уровень качества какого-либо другого изображения или какой-либо другой последовательности изображений, как это делается при некоторых видах методики оценки величин). В других случаях это число выражает суждение об уровне по сравнению с уровнем предварительно показанного «стандарта» (например, при оценке величин, разделении на группы и оценке отношений).

Результатом использования обеих видов оценок является распределение чисел, полученных для каждого испытательного условия. Используемый метод анализа зависит от типа суждения и требуемой информации (например, рангов, центральной тенденции, психологических «расстояний»).

**4.1.5.3 Методики на основе качественных показателей**

Некоторые аспекты обычного наблюдения можно выразить в терминах качественных показателей решения задач, поставленных извне (нахождение целевой информации, чтение текста, идентификация объектов и т. д.). В этом случае меру качественного показателя, такую как точность или скорость, с которой выполнены поставленные задачи, можно использовать как показатель качества изображения или последовательности изображения.

При применении методики на основе качественных показателей дается распределение оценок точности или скорости для каждого испытательного условия. При анализе основное внимание уделяется установлению соотношений между условиями применительно к центральной тенденции (и дисперсии) оценок и часто требуется применение дисперсионного анализа или аналогичного метода.

**4.1.6 Проблемы****4.1.6.1 Диапазон испытательных условий и фиксирование оценок**

Поскольку методика оценки категорий и некоторые методики без оценки категорий чувствительны к изменению диапазона и распределения рассматриваемых испытательных условий, в сеансах с вынесением суждений должны использоваться полные диапазоны изменения факторов. Однако приблизительно такой же ситуации можно достичь и при более ограниченном диапазоне, если предъявить также некоторые условия, которые соответствуют крайним точкам шкалы. Они могут быть предъявлены в виде примеров и указаны как крайние (прямая фиксация) или распределены в пределах сеанса без указания на то, что они являются крайними (косвенная фиксация).

#### 4.1.6.2 Значение оценок

Поскольку оценки меняются при изменении диапазона условий, может оказаться неправомерной интерпретация суждений, полученных с использованием методики оценки категорий и некоторых методик без оценки категорий в абсолютных понятиях (например, качество изображения или последовательность изображений).

### 4.2 Методики на основе сравнения стимулов

Согласно методикам на основе сравнения стимулов предъявляются два изображения или две последовательности изображений, а наблюдатель определяет показатель соответствия между этими двумя предъявлениями.

#### 4.2.1 Наблюдатели

Выбор наблюдателей осуществляется таким же образом, как и в случае одностимульных методик.

#### 4.2.2 Испытательные изображения

Используемые изображения или последовательности изображений формируются таким же образом, как и в случае одностимульных методик. Полученные изображения или последовательности затем комбинируются попарно, а пары используются в оценочных испытаниях.

#### 4.2.3 Условия наблюдения

Условия наблюдения выбираются таким же образом, как и при одностимульных методиках.

#### 4.2.4 Испытательный сеанс

В оценочном испытании должны использоваться либо один видеомонитор либо два хорошо согласованных видеомонитора; такое испытание, как правило, проводится так же, как и в случае одностимульных методик. При использовании одного видеомонитора в испытание включается поле дополнительного стимула, предъявляемого с той же продолжительностью, что и первого. В этом случае правильный подход состоит в том, чтобы за время всех испытаний оба элемента пары предъявлялись одинаково часто как в первой, так и во второй позиции. В случае использования двух видеомониторов оба стимульных поля предъявляются одновременно.

#### 4.2.5 Виды методик на основе сравнения стимулов

В телевизионных субъективных испытаниях используются три вида методик на основе сравнения стимулов.

##### 4.2.5.1 Методики оценки категорий

При применении методик оценки категорий наблюдатели устанавливают соответствие между элементами любой пары и системой категорий, которые обычно определяются семантически. Эти категории могут указывать на наличие наблюдаемых различий (например, **ОДИНАКОВЫ**, **РАЗЛИЧНЫ**), наличие и направление наблюдаемых различий (например, **МЕНЬШЕ**, **ОДИНАКОВО**, **БОЛЬШЕ**) или суждения о степени и направлении. Сравнительная шкала МККР приведена в таблице 5, ниже.

ТАБЛИЦА 5

Сравнительная шкала

-3	Намного хуже
-2	Хуже
-1	Слегка хуже
0	Одинаково
+1	Слегка лучше
+2	Лучше
+3	Намного лучше

Данный метод дает распределение суждений в пределах всех категорий шкалы для каждой пары испытательных условий. Выбор способа анализа результатов зависит от вынесенного суждения (например, различия) и требуемой информации (например, едва заметные различия, ранги условий, «расстояния» между условиями и т. д.).

#### **4.2.5.2 Методики без оценки категорий**

При вынесении суждений без оценки категорий наблюдатели определяют значение для степени соответствия между элементами оцениваемой пары. Имеются две разновидности этой методики:

- Согласно методике на основе непрерывной шкалы наблюдатель определяет для каждого соответствия точку на линии, проведенной между двумя словесными отметками (например, **ОДИНАКОВЫ—РАЗЛИЧНЫ** или концами такой шкалы категорий, как указано в таблице 5). Шкалы могут иметь дополнительные словесные отметки в промежуточных точках. Расстояние от одного из концов линии принимается как значение для каждой пары условий.
- Согласно второму виду методики наблюдатель определяет для каждого соответствия некое число, которое отражает оцененный уровень специфицированного параметра (например, различие в качестве). Диапазон этих чисел может быть как ограниченным, так и неограниченным. Полученное число может описывать соответствие либо в «абсолютных» понятиях, либо с точки зрения соответствия в «стандартной» паре.

Обе разновидности методик дают распределение значений для каждой пары испытательных условий. Метод анализа зависит от характера суждения и требуемой информации.

#### **4.2.6 Методики на основе качественных показателей**

В некоторых случаях критерии качественных показателей можно определить по методике сравнения стимулов. Согласно методике принудительного выбора пара формируется таким образом, что один из элементов пары обладает определенным уровнем какого-либо атрибута (например, ухудшения), а другой элемент либо обладает другим уровнем, либо вообще не обладает таким атрибутом. Наблюдателя просят определить, какой из элементов обладает большим/меньшим уровнем данного атрибута или какой из них вообще обладает этим атрибутом; меры точности и скорости принимаются как признаки соответствия между элементами пары.

#### **4.2.7 Проблемы**

##### **4.2.7.1 Формирование пар**

Методики на основе сравнения стимулов позволяют более полно оценивать соответствия между испытательными условиями в том случае, когда суждения охватывают сравнения всех возможных пар условий. Однако, если это требует слишком большого числа предъявлений, то можно либо произвести распределение предъявлений среди наблюдателей, либо прибегнуть к использованию выборок из всех возможных пар.

##### **4.2.7.2 Методики на основе многомерных шкал оценок**

Для рассмотрения суждений о сравнении стимулов в телевидении некоторые исследователи применяли методики на основе многомерных шкал оценок.

#### **4.3 Выбор методик**

Все описанные до сих пор методики имеют как сильные, так и слабые стороны, и пока нет возможности рекомендовать использовать только одну из них. Таким образом, выбор методики, наиболее подходящий для конкретного случая, остается за исследователем.

Ограничения, присущие тем или иным методикам, наталкивают на мысль, что неразумно слишком полагаться на какую-либо одну методику. Таким образом, было бы целесообразно рассмотреть более «завершенные» подходы, такие как совместное использование нескольких методик или применение многомерной методики.

## ДОПОЛНЕНИЕ 1 К ПРИЛОЖЕНИЮ 1

### Характеристики деградации изображения, обусловленной содержанием

#### 1. Введение

Реализованная система потенциально будет иметь дело с большим разнообразием программного материала, часть которого она не сможет передать без потери качества. При обсуждении пригодности конкретной системы необходимо знать пропорцию программного материала, который окажется критичным для нее, и ожидаемую при этом потерю качества. По существу, для рассматриваемой системы требуется знать характеристику деградации изображения, обусловленной содержанием.

Такая характеристика особенно важна для систем, качественные показатели которых не деградируют монотонно по мере того, как программный материал становится все более критичным. Например, некоторые цифровые и адаптивные системы могут сохранять высокое качество в широком диапазоне программного материала, однако деградируют вне этого диапазона.

#### 2. Построение характеристики деградации

С концептуальной точки зрения в характеристике деградации изображения, обусловленной содержанием, указана пропорция программного материала, который обладает большой вероятностью встретиться в течение длительного времени, для которого система обеспечивает определенный уровень качества. Это показано на рис. 5.

РИСУНОК 5

Графическое представление возможной характеристики деградации изображения, обусловленной содержанием

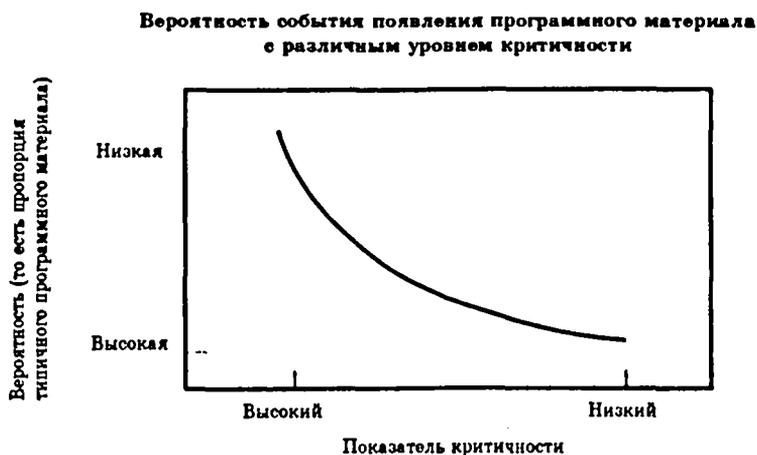


Характеристику деградации изображения, обусловленной содержанием, можно определить в четыре этапа:

**Этап 1** предусматривает определение алгоритмической меры «критичности» материала, которая должна позволять ранжирование ряда последовательностей изображений, подвергающихся искажениям в рассматриваемой системе или системах такого класса, причем таким образом, что порядок ранжирования соответствует тому порядку ранжирования, который был бы определен в случае выполнения этой задачи наблюдателем-человеком. Мера критичности может охватывать некоторые аспекты визуального моделирования.

**Этап 2** предусматривает нахождение — посредством применения меры критичности к большому числу выборок из типичных телевизионных программ — распределения, которое дает оценку вероятности события появления материала, представляющего различные уровни критичности для рассматриваемой системы или класса систем. Пример такого распределения показан на рис.6.

РИСУНОК 6



Этап 3 предусматривает определение эмпирическим путем способности системы сохранять качество по мере повышения уровня критичности программного материала. На практике для этого требуется субъективная оценка качества, обеспечиваемого данной системой для материала, отобранного таким образом, чтобы он представлял диапазон уровней критичности, определенный на этапе 2. Результатом этого является функция зависимости качества, обеспечиваемого системой, от уровня критичности программного материала. Пример такой функции приведен на рис. 7.

РИСУНОК 7

**Возможная функция зависимости качества от уровня  
критичности программного материала**



Этап 4 предусматривает объединение информации, полученной на этапах 2 и 3, для построения характеристики деградации изображения, обусловленной содержанием, в представленном на рис.5, виде.

### 3. Использование характеристики деградации

В получении полного представления о качественных показателях системы, которые она может обеспечить для всего разнообразия программных материалов, характеристика деградации играет важную роль как инструмент для оценки пригодности системы. Характеристику деградации можно использовать тремя способами:

- для оптимизации параметров системы (например, разрешающей способности источника, цифрового потока, ширины полосы частот) на стадии проектирования с целью ее более полного согласования с требованиями службы;
- для рассмотрения пригодности какой-либо одной системы (то есть для прогнозирования вероятности появления и степени деградации во время эксплуатации);
- для оценки относительной пригодности нескольких различных систем (то есть для сравнения характеристик деградации и определения того, какая из систем наиболее пригодна для использования). Следует отметить, что если различающиеся системы одного типа могут иметь один и тот же показатель критичности, то системы различных типов, возможно, могут пользоваться различными показателями критичности. Однако характеристика деградации выражает лишь вероятность того, что на практике будут встречаться различные уровни качества, поэтому такие характеристики могут сравниваться непосредственно, даже если они получены для различных зависимых от системы показателей критичности.

Хотя методика, описанная в настоящей Рекомендации, предусматривает средства измерения в системе характеристики деградации изображения, обусловленной содержанием, она может и не предсказать в полной мере приемлемость данной системы для телезрителя. Чтобы получить такую информацию, может оказаться необходимым просмотр программ, закодированных с использованием рассматриваемой системы, многими наблюдателями и изучение их комментариев.

## ДОПОЛНЕНИЕ 2

### К ПРИЛОЖЕНИЮ 1

#### Методика определения комбинированной характеристики деградации, обусловленной содержанием программ и условиями передачи

#### 1. Введение

Комбинированная характеристика деградации связывает воспринимаемое качество изображения с вероятностью практического события в том смысле, что явным образом учитывает как содержание программ, так и условия передачи.

Такую характеристику в принципе можно составить на основе субъективного исследования, которое предусматривает достаточное число наблюдателей, моментов проведения испытаний и точек приема для получения выборки, представляющей совокупность возможного содержания программ и условий передачи. Однако такого рода эксперимент может оказаться неосуществимым на практике.

В настоящем дополнении описывается альтернативная и легче реализуемая методика определения комбинированных характеристик деградации, предусматривающая три этапа:

- анализ содержания программ;
- анализ каналов передачи;
- получение комбинированных характеристик деградации.

#### 2. Анализ содержания программ

На этом этапе производятся две операции. Во-первых, находится подходящая мера содержания программы, во-вторых, оцениваются вероятности, с которыми значения этой меры встречаются на практике.

Мера содержания программы является статистической величиной, охватывающей такие аспекты содержания программы, которые подчеркивают способность рассматриваемой системы (систем) обеспечивать точное по восприятию воспроизведение программногo материала. Ясно, что было бы полезно, если бы эта мера основывалась на соответствующей модели восприятия. Однако при отсутствии такой модели может оказаться достаточной мера, которая охватывала бы некоторый аспект, касающийся степени разброса пространственных характеристик как в пределах телевизионных кадров/полей, так и между соседними кадрами/полями, при условии что между этой мерой и воспринимаемым качеством изображения существует приблизительно монотонная зависимость. Для систем (или классов систем), которые используют коренным образом отличающиеся подходы к предъявлению изображения, вероятно, потребуются другие меры.

После выбора подходящей меры необходимо оценить вероятности, с которыми появляются возможные значения этой статистической величины. Это может быть сделано двумя способами:

- эмпирическая процедура предусматривает анализ случайной выборки, содержащей, например, 200 программных фрагментов длительностью 10 с каждый, которые имеют студийный стандарт с разрешающей способностью, частотой кадров и форматом кадра, соответствующими рассматриваемой системе (системам). Анализ такой выборки дает относительные частоты события для значений статистической величины, которые принимаются за оценки вероятности практического события; или
- теоретический метод предусматривает использование теоретической модели для оценки вероятностей. Необходимо отметить, что, хотя эмпирический метод является предпочтительным, в некоторых конкретных случаях может возникнуть необходимость использовать теоретический метод (например, когда нет достаточной информации о содержании программ, как это имеет место при появлении новых технологий производства программ).

В результате применения описанных выше методов анализа находится распределение вероятностей значений статистической величины — содержания (см. также дополнение 1). Это распределение должно быть объединено с результатами анализа условий передачи для подготовки к заключительной стадии процедуры.

### 3. Анализ каналов передачи

На этом этапе также производятся две операции. Во-первых, определяется мера качественных показателей канала передачи; во-вторых, оцениваются вероятности, с которыми значения этой меры встречаются на практике.

Мера качественных показателей канала передачи является статистической величиной, охватывающей такие аспекты качественных показателей канала, которые влияют на способность рассматриваемой системы (систем) обеспечивать точное по восприятию воспроизведение программного материала источника. Ясно, что было бы полезно, если бы эта мера основывалась на соответствующей модели восприятия. Однако при отсутствии такой модели может оказаться достаточной мера, которая охватывала бы некоторый аспект, касающийся канала, при условии, что между этой мерой и воспринимаемым качеством изображения существует приблизительно монотонная зависимость. Для систем (или классов систем), в которых используются коренным образом отличающиеся подходы к каналному кодированию, вероятно, потребуются другие меры.

После выбора подходящей меры необходимо оценить вероятности, с которыми появляются возможные значения этой статистической величины. Это может быть сделано двумя способами:

- эмпирическая процедура, предусматривающая измерение качественных показателей канала, например, для 200 случайно выбранных моментов проведения испытаний и точек приема. Анализ такой выборки дает относительные частоты события для значений статистической величины, которые принимаются за оценки вероятности практического события; или
- теоретический метод, предусматривающий использование теоретической модели для оценки вероятностей. Необходимо отметить, что, хотя эмпирический метод является предпочтительным, в некоторых конкретных случаях может возникнуть необходимость использовать теоретический метод (например, при отсутствии достаточного объема соответствующей информации о качественных показателях канала, как это имеет место при появлении новых технологий производства программ).

В результате применения описанных выше методов анализа можно получить распределение вероятностей значений статистической величины — качественных показателей канала. Это распределение должно быть объединено с результатами анализа содержания программ для подготовки к заключительной стадии процедуры.

### 4. Нахождение комбинированной характеристики деградации

На данном этапе проводятся субъективные испытания, в ходе которых совместно варьируются содержание программ и условия передачи в соответствии с вероятностями, определенными на первых двух этапах.

В качестве базовой используется двухстимульная методика с непрерывной шкалой качества и, в частности, ее вариант с 10-секундными предъявлениями, рекомендованный для последовательностей движущихся изображений (см. приложение 1, § 3). Здесь в качестве эталона используется изображение студийного качества, сформированное в соответствующем стандарте [например, изображение с разрешающей способностью, частотой кадров и форматом кадра, удовлетворяющими требованиям рассматриваемой системы (систем)]. Напротив, во время испытания предъявляется то же самое изображение, но в таком виде, в каком оно было бы принято в рассматриваемой системе (системах) при выбранных условиях канала.

Испытательный материал и условия канала выбираются в соответствии с вероятностями, определенными на первых двух этапах осуществления данной методики. Фрагменты испытательного материала, причем каждый из них был проанализирован с целью нахождения его преобладающего значения согласно статистике содержания, составляют объединение выбора. Затем производятся выборки материала из этого объединения, с тем чтобы они охватывали весь диапазон возможных значений статистической величины, причем более редких при менее критичных уровнях и более частых при более критичных уровнях. Подобным образом выбираются возможные значения статистической величины — условий канала. Затем эти два независимых источника влияния объединяются случайным образом для получения комбинации содержания программ и условий канала, имеющих известную вероятность.

Результаты таких испытаний, связывающих воспринимаемое качество изображения с вероятностью практического события, используются затем для оценки пригодности какой-либо системы или для сравнения систем с точки зрения их пригодности.

## ПРИЛОЖЕНИЕ 2

## Методики оценки качества изображения применительно к ухудшениям, вызываемым цифровым кодированием телевизионных сигналов

## 1. Введение

Методики субъективной оценки качества и ухудшений изображения телевидения обычной четкости приведены в приложении 1, а в отношении ТВЧ — в Рекомендации 710. Применение этих методик для оценки телевизионных кодеков рассматривается в настоящем приложении.

В последнее время накоплен большой опыт оценки качественных показателей кодеков высокого качества, предназначенных для телевидения с разделным кодированием по стандарту 4:2:2 с цифровыми потоками 34, 45 и 140 Мбит/с. В этих испытаниях качественные показатели кодеков рассматривались с точки зрения базового качества декодированного изображения, качества такого изображения после выполнения студийной последующей обработки (квазилинейной цветовой рирпроекции и замедленного воспроизведения) и ухудшений декодированного изображения для различных значений коэффициента ошибок в канале. В ряде разделов настоящего приложения описывается опыт такого рода.

В случае применений, связанных с распределением программ, качество можно специфицировать в терминах субъективных суждений наблюдателей. Следовательно, такие кодеки теоретически могут оцениваться субъективно со ссылкой на указанные спецификации. Однако качество кодека, разработанного для применений, связанных со снабжением программ, теоретически не может специфицироваться в терминах субъективной оценки качественных параметров, поскольку его выходной сигнал предназначен не для непосредственного просмотра, а для студийной постобработки, запоминания и/или кодирования для дальнейшей передачи. Из-за трудностей определения этих качественных показателей для множества различных видов последующей обработки предпочтительным подходом является спецификация качественных показателей всего тракта с оборудованием, включая операцию постобработки; считается, что этот подход представляет собой один из вариантов практического применения по снабжению программами. Такой тракт в типичном случае может содержать кодек, за которым следует аппаратура студийной постобработки (или другой кодек в случае оценки базового качества при снабжении программами), и еще один кодек, с выхода которого сигнал поступает к наблюдателю. Принятие такой стратегии для спецификации кодеков, предназначенных для применений по снабжению программами, означает, что процедуры испытаний, приведенные в настоящей Рекомендации, могут также использоваться для их оценки.

В настоящем приложении подчеркивается важность выбора критичных испытательных последовательностей изображений, причем в основном реальных сюжетов, а также даны указания в отношении того, как такие последовательности могут быть созданы или выбраны.

## 2. Субъективная оценка качества изображения при использовании кодеков

Несмотря на определенный прогресс, в настоящее время еще не накоплен достаточный опыт, позволяющий дать подробную разработку методик объективной оценки качества изображения применительно к кодекам. В области субъективной оценки, где накоплен большой опыт, условия испытаний и методики могут быть рекомендованы. Однако при спецификации целевых значений качества или ухудшений изображения следует помнить, что существующие методики не дают абсолютных субъективных оценок, а скорее результаты, на которые в некоторой степени влияет выбор эталонных изображений и/или условий фиксирования оценок. Одни и те же методики могут быть приняты для кодеков как с фиксированной, так и с переменной длиной слова, а также для внутрисловых и межкадровых кодеков, хотя выбор испытательных последовательностей изображений может быть различным.

В настоящее время самой надежной методикой оценки порядка ранжирования высококачественных кодеков является оценка всех предлагаемых систем в одно и то же время и в идентичных условиях. В тех случаях, когда различия в качестве незначительны, независимо проводимые испытания следует воспринимать скорее как руководство, а не как неоспоримое доказательство превосходства.

## 2.1 Оценка базового качества

Если оценивается кодек, используемый для распределения программ, то имеется в виду качество декодированных изображений, полученных после однократного прохождения через такой кодек. В случае кодеков, используемых для снабжения программами, базовое качество может оцениваться после прохождения сигнала через несколько кодеков, включенных последовательно, поскольку это моделирует типичное применение по снабжению программами.

### **2.1.1 Условия наблюдения и выбор наблюдателей**

Рекомендуется, чтобы условия наблюдения и выбор наблюдателей были такими же, какие описаны в § 2.4 приложения 1 в отношении телевидения обычной четкости, а также в Рекомендации 710 в отношении кодеков ТВЧ.

### **2.1.2 Использование испытательных последовательностей изображений**

Рекомендуется, чтобы в испытаниях использовались по крайней мере шесть последовательностей изображений и одна дополнительная последовательность для демонстрационных целей до начала сеансов. Требуется использовать последовательности продолжительностью порядка 10 с, однако следует заметить, что наблюдатели могут предпочесть последовательности продолжительностью 15—30 с. Последовательности, используемые в контексте рассматриваемого сокращения цифрового потока, должны быть умеренно критичными или критичными.

### **2.1.3 Методика испытаний**

Если оцениваемый диапазон качества ограничен, что обычно бывает в случае телевизионных кодеков, то применяется двухстимульная методика с непрерывной шкалой качества, описанная в § 3 приложения 1. За эталон принимается последовательность исходных изображений источника. Проводятся дальнейшие исследования в отношении продолжительности предъявления последовательностей. В испытаниях кодеков телевизионного стандарта 4:2:2 с отдельным кодированием, проведенных в последнее время, было признано целесообразным модифицировать способ предъявления, приведенный в настоящей Рекомендации. В качестве дополнительного эталона были использованы изображения совместного кодирования, обеспечивавшие пониженный уровень качества, относительно которого выносились суждения о качественных показателях кодеков.

## **2.2 Оценка качества в результате постобработки**

Такое испытание позволяет вынести суждение о пригодности кодека для применения связанных со снабжением программами с точки зрения конкретных операций постобработки, таких как квазилинейная цветовая рирпроекция, замедленное воспроизведение, электронное масштабирование изображения. Минимальный набор аппаратуры для такого испытания должен обеспечивать однократное прохождение сигнала через испытываемый кодек, затем выполнение требуемой постобработки и наблюдение. Однако более представительным для данного применения по снабжению программами было бы включение дополнительных кодеков после аппаратуры постобработки.

### **2.2.1 Условия наблюдения и выбор наблюдателей**

См. § 2.1.1.

### **2.2.2 Использование испытательных последовательностей изображений**

Вследствие практических ограничений, связанных с возможным требованием оценки кодека в сочетании с несколькими операциями постобработки, число испытательных последовательностей изображений должно быть не менее трех с одной добавочной последовательностью, используемой для демонстрационных целей. Характер таких последовательностей зависит от исследуемой операции постобработки, однако в контексте сокращения цифрового потока и с учетом характера рассматриваемой постобработки они должны быть умеренно критичными или критичными. Требуется использовать последовательности продолжительностью порядка 10 с, однако следует заметить, что наблюдатели могут предпочесть последовательности продолжительностью 15—30 с. Для оценки замедленного воспроизведения может оказаться приемлемой частота отображения, составляющая 1/10 частоты отображения источника.

### **2.2.3 Методика испытаний**

Для испытаний применяется двухстимульная методика с непрерывной шкалой качества. Однако в данном случае эталоном служит изображение источника, подвергнутое такой же постобработке, что и декодированное изображение. Если решено, что полезно включить эталон более низкого качества, то он также должен быть подвергнут той же самой постобработке. В испытаниях, проведенных в МККР, использовался несколько измененный способ предъявления по сравнению с тем, который описан в настоящей Рекомендации.

## **3. Субъективная оценка ухудшений изображений, вызываемых кодеком, с учетом ошибок передачи**

Полезной субъективной мерой может быть ухудшение, определенное как функция коэффициента ошибок, которая имеет место в линии передачи между кодером и декодером. В настоящее время накоплено недостаточно экспериментальных данных по реальной статистике ошибок передачи, чтобы можно было рекомендовать параметры для модели, учитывающей кластерные или пакетные ошибки. До сбора такой информации для ошибок можно пользоваться распределением Пуассона.

### 3.1 *Использование испытательных последовательностей изображений*

Поскольку требуется исследовать качественные показатели кодека в некоем диапазоне коэффициентов ошибок при передаче, с учетом практических ограничений можно предположить, что, вероятно, будет достаточно использовать три испытательных последовательности изображений и одну дополнительную демонстрационную последовательность. Каждая последовательность должна иметь продолжительность порядка 10 с, однако следует заметить, что наблюдатели могут предпочесть последовательности продолжительностью 15—30 с. В контексте сокращения цифрового потока в телевидении последовательности должны быть умеренно критичными или критичными.

### 3.2 *Выбор коэффициентов ошибок*

Должно быть выбрано не менее пяти, а предпочтительно больше, коэффициентов ошибок с расположением согласно логарифмическому закону в диапазоне, который обеспечивает изменение ухудшений, вызываемых кодеком, в пределах от «незаметно» до «очень мешает».

### 3.3 *Методика испытаний*

Поскольку в испытаниях используется весь диапазон возможных ухудшений, двухстимульная методика со шкалой ухудшений является пригодной, и ее следует применять.

### 3.4 *Замечание об использовании очень малых коэффициентов ошибок*

Может потребоваться оценка кодеков при таких значениях коэффициента ошибок передачи, что видимые переходные явления будут наблюдаться очень редко; их появления можно даже и не ждать в течение испытательной последовательности продолжительностью 10 с. Предложенный здесь временной порядок предъявления явно не подходит для таких испытаний.

Может производиться видеозапись кодека изображения на выходе при условии довольно малого коэффициента ошибок (приводящего к малому числу видимых переходных явлений в пределах периода 10 с), причем полученные видеофонограммы будут в дальнейшем монтироваться, чтобы их можно было использовать для предъявления в сеансах субъективной оценки. К этому следует относиться осторожно, чтобы гарантировать, что используемые видеофонограммы являются типичными для выходного изображения кодека, наблюдаемого в течение довольно длительного периода времени.

## 4. *Субъективные сравнения кодеков*

В тех случаях, когда не требуется выносить суждений об абсолютном качестве кодека или ухудшении, а требуется определить только порядок ранжирования, или в том случае, когда необходимо подтвердить порядок ранжирования, установленный на основе результатов применения двухстимульной методики, следует пользоваться методикой сравнения парных стимулов.

Как уже говорилось, эта методика обеспечивает «чувствительное» сравнение и средство определения меры соответствия между парами систем. Возможно расширение применимости этой методики для ранжирования степеней качества или ухудшения, когда число систем более двух. При таком подходе результирующий порядок ранжирования определяется на основе порядка ранжирования всех возможных пар последовательностей изображений, определенных наблюдателями.

Анализ осложняется тем, что наблюдатель может, например, установить ранг изображения А выше ранга изображения В, ранг изображения В — выше ранга изображения С и одновременно ранг изображения С — выше ранга изображения А. Это называется «интранзитивной триадой».

Трудности методики заключаются в том, что число требуемых предъявлений возрастает как квадрат числа испытательных последовательностей изображений и кодеков; реализация такого числа может оказаться практически невозможной.

## 5. *Выбор испытательных изображений для оценки цифровых кодеков*

В настоящем приложении подчеркивается важное значение испытаний цифровых кодеков с использованием последовательностей изображений, которые критичны в контексте сокращения цифрового потока в телевидении. Поэтому уместно задаться вопросом, насколько критичной является конкретная последовательность изображений для конкретной задачи сокращения цифрового потока, или является ли одна последовательность более критичной, чем другая. Простой, но не очень полезный ответ заключается в том, что «критичность» означает весьма различные вещи для разных кодеков. Например, для внутривидеопотокового кодека может быть довольно критичным неподвижное изображение, содержащее много деталей, а для межкадрового кодека, который способен использовать сходство соседних кадров, такое же изображение не вызовет никаких трудностей. Некоторые последовательности, включающие подвижные текстуры и сложное движение, будут критичными для всех классов кодеков; именно поэтому создание или распознавание последовательностей такого рода наиболее полезно. Сложное движение может принимать форму таких перемещений, которые прогнозируются наблюдателем, но не алгоритмом кодирования; примером может служить «рваное» периодическое движение.

При одном из исследований возможных статистических мер критичности изображения, определенных с помощью корреляционных методов, спектральных методов, методов условной энтропии и т.д., обнаружилась простая, но очень удобная мера, основанная на адаптивном измерении внутрисюжетной/межсюжетной энтропии. Этот метод использовался для «калибровки» последовательностей изображений, предложенных в МККР для испытания кодеков с потоком 34, 45 и 140 Мбит/с, и оказался полезным для выбора используемых последовательностей. Проведение таких измерений параметров последовательностей изображений легче всего осуществить посредством загрузки их в компьютеры для обработки изображений и проведения их дальнейшего анализа с помощью программного обеспечения.

Если такие методы недоступны, то ниже приводятся некоторые общие рекомендации по выбору критичного материала.

**а) Внутрисюжетные кодеки с фиксированной длиной слова**

Хотя возможно и корректно оценивать такие кодеки с использованием неподвижных изображений, рекомендуется пользоваться последовательностями движущихся изображений, поскольку при этом проще наблюдать шумовые процессы кодирования, а кроме того, такие последовательности более свойственны телевизионным применениям. Если в компьютерном имитационном моделировании кодеков используются неподвижные изображения, то обработка должна выполняться по всей оцениваемой последовательности, чтобы, например, полностью учесть временные аспекты всех источников шума. Выбранные сюжеты должны содержать как можно больше следующих деталей: области с неподвижными или движущимися текстурами (некоторые из них с цветной текстурой); неподвижные и движущиеся объекты с резкими, высококонтрастными границами различной ориентации (некоторые из них цветные); неподвижные однородные серые области средней яркости. По крайней мере одна последовательность в ансамбле должна содержать едва заметные шумы источника и по крайней мере еще одна должна быть синтезированной (то есть генерирована компьютером), для того чтобы в ней отсутствовали дефекты телекамеры, такие как апертурные искажения и инерционность.

**б) Межсюжетные кодеки с фиксированной длиной слова**

Все выбранные испытательные сюжеты должны содержать движущиеся изображения и как можно больше следующих деталей: области с движущейся текстурой (некоторые из них с цветной); объекты с резкими высококонтрастными границами, движущимися в направлении, перпендикулярном к самим себе, и имеющими различную ориентацию (некоторые из них цветные). По крайней мере одна последовательность в ансамбле должна содержать едва заметные шумы источника и по крайней мере еще одна должна быть синтезированной.

**в) Внутрисюжетные кодеки с переменной длиной слова**

Рекомендуется испытывать такие кодеки с помощью материала, содержащего последовательности движущихся изображений, по тем же причинам, что и в случае кодека с фиксированной длиной слова. Следует отметить, что благодаря достоинствам кодирования с переменной длиной слова и наличию соответствующей буферной памяти эти кодеки могут динамически распределять «битовую емкость» кодирования по всему изображению. Так, например, если половина изображения представляет собой чистое небо, для кодирования которого не требуется большого числа битов, битовая емкость резервируется для других частей изображения, которые благодаря этому могут быть отображены с высоким качеством, даже если они критичны. Из вышесказанного можно сделать важный вывод, что если необходимо иметь последовательность изображений, критичную для таких кодеков, то содержание изображения в каждой части экрана должно иметь множество деталей. Изображение должно быть заполнено движущимися и неподвижными текстурами с как можно большими цветовыми вариациями и объектами с резкими, высококонтрастными границами. По крайней мере одна последовательность в ансамбле текстов должна содержать едва заметные шумы источника и по крайней мере еще одна последовательность должна быть синтезированной.

**г) Межсюжетные кодеки с переменной длиной слова**

Это наиболее совершенный класс кодеков, причем для полной загрузки этой разновидности кодеков требуется наиболее трудный материал. Каждая часть сюжета должна быть заполнена деталями, как в случае внутрисюжетных кодеков с переменной длиной слова, но и сами детали должны находиться в движении. Более того, поскольку во многих кодеках используются методы компенсации движения, во всей испытательной последовательности движение должно быть сложным. Примеры сложного движения: сюжеты с одновременным масштабированием изображения и панорамированием телекамеры; сюжеты, в которых в качестве фона используется занавес с текстурой или мелкими деталями, развевающийся на ветру; сюжеты, содержащие объекты, вращающиеся в трех измерениях; сюжеты, содержащие объекты с высокой детальностью, движущиеся с ускорением по экрану. Все сюжеты должны предусматривать движение объектов, имеющих различные скорости, текстуры и высококонтрастные границы, а также варьирование содержания цветов. По крайней мере одна последовательность в ансамбле текстов должна содержать едва заметные шумы источника и еще одна последовательность — имитацию сложного движения камеры, созданную с помощью компьютера на основе реального неподвижного изображения (для того, чтобы в ней отсутствовали шум и инерционность камеры), и по крайней мере еще одна последовательность должна быть полностью генерирована компьютером.

Испытательные последовательности, требуемые для оценки постобработки, должны соответствовать точно таким же критериям критичности. Однако это может оказаться труднодостижимым в случае формирования последовательностей изображений переднего плана для цветовой рирпроекции, поскольку обычно они имеют значительную пропорцию гладкого синего фона.

Обширная библиотека испытательных последовательностей была подготовлена с использованием стандарта 4:2:2, причем она хранится в виде видеофонограмм стандарта D1. Подробные сведения относительно этих последовательностей, а также критерии, на основе которых они были подготовлены (и которые могут быть применены к другим видеостандартам), приводятся в Рекомендации 802.

---