

МСЭ-Т

СЕКТОР СТАНДАРТИЗАЦИИ
ЭЛЕКТРОСВЯЗИ МСЭ

P.800.2

(07/2016)

СЕРИЯ Р: ОКОНЕЧНОЕ ОБОРУДОВАНИЕ,
СУБЪЕКТИВНЫЕ И ОБЪЕКТИВНЫЕ
МЕТОДЫ ОЦЕНКИ

Методы для объективной и субъективной оценки
качества речи и видео

**Интерпретация и представление средней
экспертной оценки**

Рекомендация МСЭ-Т P.800.2

РЕКОМЕНДАЦИЯ МСЭ-Т СЕРИИ Р

ОКОНЕЧНОЕ ОБОРУДОВАНИЕ, СУБЪЕКТИВНЫЕ И ОБЪЕКТИВНЫЕ МЕТОДЫ

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
Характеристики 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.800.2

Интерпретация и представление средней экспертной оценки

Резюме

В Рекомендации МСЭ-Т Р.800.2 вводится ряд новых общих типов средней экспертной оценки (MOS) и определяется минимальная информация, которая должна сопровождать значения MOS для гарантирования их корректной интерпретации.

Хронологическая справка

Издание	Рекомендация	Утверждение	Исследовательская комиссия	Уникальный идентификатор*
1.0	МСЭ-Т Р.800.2	14.05.2013 год	12-я	11.1002/1000/11934
2.0	МСЭ-Т Р.800.2	29.07.2016 год	12-я	11.1002/1000/12973

Ключевые слова

Оценка в абсолютных категориях, ACR, средняя экспертная оценка, MOS, объективная модель, представление, субъективный эксперимент.

* Для получения доступа к Рекомендации наберите в адресном поле вашего браузера URL-адрес <http://handle.itu.int/>, а затем 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", "должен" (shall) или некоторые другие обязывающие выражения, такие как "обязан" (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 извещения 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 2019

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 применение	1
2 Справочные документы	1
3 Определения	1
3.1 Термины, определяемые в других документах	1
3.2 Термины, определяемые в настоящей Рекомендации	1
4 Сокращения и акронимы	1
5 Соглашения по терминологии	2
6 Вводная информация	2
7 Субъективные значения MOS	2
8 Интерпретация значений MOS	4
9 Соображения, касающиеся видеосигналов	5
10 Статистический анализ MOS	5
11 Объективные значения MOS	5
12 Представление субъективных значений MOS	6
13 Представление объективных значений MOS	8
14 Условные обозначения	8
Библиография	9

Рекомендация МСЭ-Т Р.800.2

Интерпретация и представление средней экспертной оценки

1 Сфера применения

В настоящей Рекомендации вводится ряд новых общих типов средней экспертной оценки (MOS) и определяется минимальная информация, которая должна сопровождать значения MOS для гарантирования их корректной интерпретации.

Следует отметить, что этот текст не имеет целью служить исчерпывающим руководством по проведению субъективных или объективных испытаний. Источники более подробной информации по этой теме перечисляются в разделе "Библиография" в конце настоящей Рекомендации.

2 Справочные документы

В нижеследующих Рекомендациях МСЭ-Т и других справочных документах содержатся положения, которые посредством ссылок на них в данном тексте составляют положения настоящей Рекомендации. На момент публикации указанные здесь издания были действующими. Все Рекомендации и другие справочные документы могут подвергаться пересмотру; поэтому всем пользователям данной Рекомендации предлагается изучить возможность применения последнего издания Рекомендаций и других справочных документов, перечисленных ниже. Список действующих в настоящее время Рекомендаций МСЭ-Т регулярно публикуется. Ссылка на документ в данной Рекомендации не придает ему как отдельному документу статус Рекомендации.

[ITU-T R.800.1] Рекомендация МСЭ-Т Р.800.1 (2006 год), *Терминология, касающаяся средней оценки разборчивости речи (MOS)*.

3 Определения

3.1 Термины, определяемые в других документах

Отсутствуют.

3.2 Термины, определяемые в настоящей Рекомендации

В настоящей Рекомендации определяются следующие термины.

3.2.1 условие (condition): одно из условий работы в наборе сценариев использования, оцениваемое в субъективном эксперименте; часто соотносят с гипотетическим эталонным трактом (HRC) в экспериментах по оценке видеоизображения.

3.2.2 субусловие (sub-condition): поднабор параметров для того или иного условия работы, определяемый конкретной характеристикой определенного сценария использования, например речь конкретного говорящего субъекта.

3.2.3 участник (subject): лицо, участвующее в субъективном эксперименте.

3.2.4 голос (vote): ответ участника на вопрос о количественной оценке отдельного испытательного образца или интерактивного диалога по установленной шкале.

4 Сокращения и акронимы

В настоящей Рекомендации используются следующие сокращения и акронимы.

ACR	Absolute Category Rating	Оценка в абсолютных категориях
DCR	Degradation Category Rating	Оценка в категориях ухудшения
DMOS	Degradation Mean Opinion Score	Средняя экспертная оценка ухудшения
HRC	Hypothetical Reference Circuit	Гипотетический эталонный тракт

MOS	Mean Opinion Score	Средняя экспертная оценка
MUSHRA	Multi-stimulus test with Hidden Reference and Anchor	Тест при использовании нескольких входных сигналов со скрытым эталонным сигналом и опорным сигналом
QCIF	Quarter Common Intermediate Format	Четвертной единый промежуточный формат
SSCQE	Single Stimulus Continuous Quality Evaluation	Метод непрерывной оценки качества при одном источнике воздействия
VGA	Video Graphics Array	Видеографическая матрица

5 Соглашения по терминологии

Отсутствуют.

6 Вводная информация

Качество аудио- и видеосигналов – величины изначально субъективные. Это означает, что критерием для их оценки служит мнение пользователя. Но мнения разных людей о том, что есть хорошо, могут сильно различаться, и никого из них нельзя назвать правым или неправым.

Прежде чем приступать к развертыванию новой технологии передачи аудио- или видеосигналов, целесообразно оценить качество передачи, организовав один или несколько субъективных экспериментов. Цель субъективного эксперимента – собрать мнения множества людей (участников) о рабочих характеристиках системы для ряда четко определенных сценариев использования (условий¹). Средняя экспертная оценка (MOS) для данного условия представляет собой просто среднее арифметическое значение оценок (голосов), полученных для соответствующего сценария использования.

Задачей алгоритмов объективного измерения качества является прогнозирование значения MOS, которое будет получено при заданном входном сигнале в субъективном эксперименте. Следовательно, при интерпретации значения MOS, полученного объективным путем, важно понимать общий план эксперимента, результаты которого прогнозируются.

Существует несколько типов значений MOS и целый ряд методов испытаний, предназначенных для их получения. Цель настоящей Рекомендации – провести анализ основных соображений, которые читателю следует принимать во внимание при интерпретации значений MOS, и предоставить минимальную информацию, которая должна сопровождать сообщаемые значения MOS.

7 Субъективные значения MOS

Типы значений MOS

Распространено заблуждение, что значения MOS относятся только к голосовым услугам, но процесс опроса участников эксперимента относительно их оценок качества можно столь же легко применить к видео- и аудиослугам общего характера. Можно также попросить участников оценить общее качество аудиовизуальных программ, обеспечиваемое услугой. В рамках МСЭ разработаны различные стандарты, в которых описываются разнообразные аспекты субъективного тестирования для видеоприменений и аудиоприменений общего характера в дополнение к голосовым применениям. Эти стандарты перечисляются в разделе "Библиография".

Субъективные эксперименты можно в общем разделить на два типа – пассивные и интерактивные. В пассивном субъективном эксперименте участникам предъявляются заранее записанные испытательные образцы, которые воспроизводят представляющие интерес условия. Участников просят

¹ В экспериментах с видеоизображениями условия часто соотносят с гипотетическими эталонными трактами (HRC).

пассивно прослушать и/или просмотреть испытательный материал и дать свою оценку по установленной шкале. В интерактивном эксперименте двое или несколько участников ведут активную беседу с использованием оборудования, предназначенного эмулировать представляющие интерес сценарии использования. Часто участникам дают те или иные задания, чтобы стимулировать разговор и взаимодействие. Большинство экспериментов носят пассивный характер. Имеются однако некоторые аспекты взаимодействия с пользователем, например влияние задержки и эха, которые проявляются только в разговорных сценариях.

Методика испытания и шкала оценок

В субъективном эксперименте участников просят выразить свое мнение по шкале оценок. Назначение этой шкалы – преобразовать оценку качества, данную участником, в числовое значение, которое можно усреднить по множеству участников и другим факторам эксперимента.

Существует несколько повсеместно принятых шкал оценок, обсуждение их относительных преимуществ выходит за рамки настоящей Рекомендации. Чаще всего применяется 5-балльная шкала оценок в абсолютных категориях (ACR):

Отлично	5
Хорошо	4
Удовлетворительно	3
Плохо	2
Неприемлемо	1

Шкала ACR дискретная; это означает, что ответы участника ограничиваются одним из пяти перечисленных выше значений. Но поскольку результаты разных участников комбинируются путем усреднения, значения MOS не обязательно являются целочисленными. Некоторые шкалы оценок имеют более пяти дискретных градаций, в то время как другие шкалы позволяют участнику давать ответы в промежутках между обозначенными градациями.

Прилагательное "абсолютный" в названии шкалы ACR указывает на то, что участников просят оценить каждый образец безотносительно к другим. При использовании некоторых шкал, например шкалы оценок в категориях ухудшения (DCR), участника просят оценить разницу в качестве между образцом, который был подвергнут обработке в соответствии с представляющим интерес условием, и необработанной версией того же образца. Значение MOS, полученное в таком эксперименте, часто называют "MOS ухудшения", или DMOS.

Большинство экспериментов планируются таким образом, что участников просят оценить качество коротких аудио- или видеообразцов. Длительность таких образцов находится диапазоне от 6 до 10 секунд, так как это дает участнику достаточно времени, чтобы он составил собственное мнение, но к моменту окончания воспроизведения образца у него еще не успела бы сформироваться систематическая ошибка оценки. Трудно в полной мере воспроизвести условие в одном образце такой длительности, поэтому участников обычно просят оценить множество испытательных образцов, отражающих один и тот же сценарий использования. Например, в голосовом эксперименте каждое из испытываемых условий функционирования сети может быть представлено образцами речи, полученными от шести говорящих субъектов (по три субъекта мужского и женского пола). Это означает, что значения MOS могут быть получены для всего условия в целом (путем усреднения по участникам и говорящим субъектам) или для субусловия – например, для конкретного говорящего субъекта или пола говорящего субъекта.

В таких методах испытаний, как метод непрерывной оценки качества при одном источнике воздействия (SSCQE), применяются испытательные образцы намного большей длительности, и от участника требуется непрерывно обновлять свою оценку качества воспроизводимого испытательного образца. Результатом является временная последовательность оценок качества от каждого участника, а не одиночное значение оценки.

В некоторых методах испытаний участника просят ответить на несколько вопросов. Это не только дает больше информации о тестируемых условиях, но и может быть необходимой составляющей плана эксперимента. Например, метод испытаний, описываемый в Рекомендации МСЭ-Т Р.835, требует от участника отдельно оценить качество речи и качество шума в предъявленном образце, прежде чем давать общую оценку качества. Установлено, что этот процесс дает более устойчивые результаты с системами шумоподавления, чем метод испытаний с одиночным вопросом по шкале ACR.

Следует отметить, что некоторые вопросы могут быть не связаны напрямую с качеством, а затрагивать другой аспект связи: например, в [b-ITU-T P.800] определяется шкала *необходимых усилий* при прослушивании для речевых экспериментов. В некоторых разговорных экспериментах участника спрашивают о его ощущениях при говорении, а не при прослушивании.

8 Интерпретация значений MOS

В приведенном ниже обсуждении основное внимание поначалу уделяется значениям MOS для голосовых сигналов, но многие тезисы, высказанные в подразделах, равным образом относятся к значениям MOS для видео-, аудиосигналов и сочетания сигналов аудио–видео. Основные отличия в случае видеосигналов описываются в следующем разделе.

Идея о том, что конкретный речевой кодек имеет определенное значение MOS, – это еще одно распространенное заблуждение. Одним из источников данного заблуждения является широкое использование моделей объективной оценки качества, обеспечивающих высокую степень повторяемости результатов. Такие модели предназначаются для прогнозирования или оценки результатов субъективных экспериментов, но для любого заданного кодека при заданной битовой скорости передачи значение MOS, полученное в субъективном эксперименте, может существенно варьироваться между экспериментами. Для этого есть ряд причин.

Во-первых, на точные значения MOS, полученные для конкретного условия в субъективном эксперименте, может влиять большое количество факторов, в том числе:

- указания, данные участникам, и формулировки шкалы мнений;
- оборудование, используемое для воспроизведения материала (телефонная трубка, гарнитура, громкоговорители);
- способ воспроизведения (моноуральный, диотический бинауральный или стереофонический);
- уровень воспроизведения;
- акустическая среда восприятия;
- подготовка участников;
- профиль участников, например возраст и опыт общения с техникой;
- межкультурные различия в интерпретации и использовании шкал оценки;
- характер речевого материала (фонетическое содержание и характеристики говорящего субъекта);
- язык (наличие/отсутствие, преобладание и важность конкретных звуков и переходов).

Во-вторых, точное значение MOS, полученное для любого заданного условия в субъективном эксперименте, зависит от качества других условий в том же эксперименте. Например, условие с применением речевого кодека ITU-T G.729 может дать оценку выше 3,9 в эксперименте с использованием шкалы ACR, если большинство прочих условий соответствует худшему качеству, чем ITU-T G.729; и наоборот, условие ITU-T G.729 может дать оценку существенно ниже 3,9, если большинство прочих условий соответствует более высокому качеству.

В-третьих, если эксперимент проводится с кодеками, работающими в условиях разной полосы частот аудиосигнала, то присутствие условий с более широкой полосой частот приведет к снижению значений MOS, полученных для условий с более узкой полосой частот аудиосигнала. Наиболее широкую полосу частот аудиосигнала в речевом эксперименте часто называют контекстом эксперимента. Например, условие с применением речевого кодека ITU-T G.711 может зачастую дать оценку выше 4,0 в узкополосном (300–3700 Гц) эксперименте с использованием шкалы ACR, но в широкополосном (50–7000 Гц) эксперименте с использованием шкалы ACR оно с большей вероятностью даст оценку в диапазоне 3,5–3,7 из-за присутствия более высококачественных широкополосных образцов.

Последние два обстоятельства отражают тот факт, что участники экспериментов часто используют шкалу оценки, принимая во внимание содержание эксперимента. Действительно, хорошо спланированные эксперименты обычно предусматривают начальный ознакомительный этап, когда участникам демонстрируются примеры ряда условий, в том числе наилучших и наихудших.

Одно из наиболее важных следствий из приведенных здесь соображений состоит в том, что не существует эффективного способа напрямую сравнить значения MOS, полученные в разных экспериментах, если эти эксперименты не были специально спланированы в расчете на сравнение результатов – и даже в таком случае данные следует проанализировать статистически, чтобы убедиться, что такое сравнение будет корректным.

9 Соображения, касающиеся видеосигналов

Многие из соображений, высказанных выше применительно к речевым субъективным экспериментам, применимы также и к экспериментам с видеосигналами. В соответствии с условиями эксперимента, которые часто называют гипотетическими эталонными трактами (HRC), обычно определяются различные сочетания видеокodeка, битовой скорости, частоты кадров и условий передачи. На точные значения MOS, полученные для конкретного условия, влияют, среди прочего, следующие факторы:

- оборудование, используемое для воспроизведения материала (технология дисплея, частота обновления информации, контрастность и т. д.);
- условия просмотра (цветность, температура и уровень освещенности);
- расстояние просмотра (обычно выражаемое через отношение расстояния просмотра к высоте дисплея);
- содержание видеоинформации.

Последний пункт особенно важен в видеоэкспериментах. Выбор тестируемого видеоматериала оказывает гораздо большее влияние на результат эксперимента, чем выбор речевого материала. Это связано с тем, что содержимое видеопоследовательности может в значительной степени определять достижимый уровень эффективности ее кодирования. Например, информационное наполнение спортивного видеоролика с быстро меняющейся картинкой намного выше, чем в случае видеоконференции с отображением на экране головы и плеч собеседника.

В видеоэкспериментах основной контекст определяется разрешением видеоизображения. Как правило, в субъективных экспериментах не смешивают разные разрешения, поэтому значения MOS в экспериментах с видеоматериалом относятся к конкретному разрешению, например 480p или 1080p. Когда же разрешения все-таки смешивают, контекст эксперимента определяется разрешением с наибольшим числом строк. В этом случае важно отметить, отображаются ли более низкие разрешения штатным образом или путем масштабирования до наивысшего разрешения в эксперименте.

10 Статистический анализ MOS

Статистический анализ субъективных значений MOS выходит за рамки настоящей Рекомендации. Однако значения MOS следует сопровождать информацией, достаточной для выполнения их базового статистического анализа, например, для расчета доверительного интервала по каждому условию. Для каждого заданного условия или субусловия эта информация включает количество голосов, их среднее арифметическое и среднеквадратичное отклонение.

11 Объективные значения MOS

Назначение модели объективной оценки качества – спрогнозировать значение MOS, которое будет получено для аудио- или видеосигнала в субъективном эксперименте. Как уже отмечалось выше, точное значение MOS, полученное в любом заданном эксперименте для конкретного кодека или тракта передачи, зависит от множества различных аспектов планирования и осуществления эксперимента. Поэтому проектировщикам модели объективной оценки приходится прогнозировать идеализированный эксперимент. Обычно это эксперимент, который проводится по конкретной методике, как правило, с использованием шкалы ACR и включает в себя сбалансированный образец искажений, с которыми приходится сталкиваться в рассматриваемой области применения.

Например, преобразование, определяемое в [b-ITU-T P.862.1], ставит в соответствие необработанным выходным данным модели объективной оценки согласно ITU-T P.862 диапазон, полученный путем усреднения результатов большого количества субъективных экспериментов по методу ACR, как описывается в [b-ITU-T P.800]. Сходное преобразование производится на выходе модели ITU-T P.863.

Одно из преимуществ модели объективной оценки состоит в том, что она дает воспроизводимые результаты, и поэтому измерения, выполненные в разное время и в разных местах, можно непосредственно сравнивать. Тем не менее следует соблюдать осмотрительность, поскольку такие факторы, как выбор тестируемого материала, а также предварительная и последующая обработка, могут внести систематическую погрешность в результаты.

По причинам, которые сейчас должны уже быть ясны читателю, разные модели объективной оценки могут давать различные прогнозируемые значения MOS для одних и тех же условий. Например, модели ITU-T P.862.1 и ITU-T P.863 не дают в точности одинаковые прогнозируемые значения MOS для речевого кодека ITU-T G.729, хотя этот кодек входит в сферу применения обеих моделей. Отчасти это связано с тем, что обучение и оптимизация моделей производились с использованием разных субъективных экспериментов. Поэтому при сравнении объективных прогнозируемых значений MOS с пороговыми значениями – например, для контроля соблюдения соглашения об уровне обслуживания или для подачи сигнализации – такие пороговые значения необходимо выбирать в контексте той модели, с помощью которой был получен прогноз.

12 Представление субъективных значений MOS

В таблице 1 описывается информация, которой необходимо сопровождать субъективные значения MOS, а также дополнительная информация, которую рекомендуется предоставлять.

Если проведен эксперимент согласно Рекомендации МСЭ, информацию о методике его проведения можно обычно предоставить путем простой ссылки на соответствующий стандарт и конкретный использовавшийся метод, указав отклонения от стандартной методики, если таковые были.

Важно всегда приводить информацию о тестируемых образцах, использовавшихся для пассивных экспериментов. В случае видеообразцов может оказаться целесообразным привести более детальную информацию – например, имеются ли в конкретных последовательностях панорамирование или смены сцены.

Таблица 1 – Минимальная информация для представления субъективных значений MOS

Информация	Тип эксперимента	Предоставление информации
Методика Пассивная или интерактивная Оценка по отдельным образцам или непрерывная оценка Абсолютная или относительная оценка образцов Указания и вопросы участникам Названия градаций шкалы оценки Дискретность или непрерывность шкалы оценки Длительность образца или Рекомендация МСЭ и метод, применявшиеся при оценке	Все	Обязательное

Таблица 1 – Минимальная информация для представления субъективных значений MOS

Информация	Тип эксперимента	Предоставление информации
План испытаний Цель эксперимента Дата и место проведения испытаний Сведения об обработке План эксперимента, например планирование по блокам Количество и длительность сеансов Число участников Профили участников, распределение по возрасту и полу Типы задействованных участников, например неопытные участники или эксперты Сведения об использовавшемся оборудовании Условия окружающей среды при представлении образцов, то есть уровень собственных шумов, уровень освещенности и т. д.	Все	Рекомендуемое
Сведения об условиях/HRC Количество условий Перечень условий Средние значения оценок/голосов по каждому условию (MOS) Среднеквадратичное отклонение голосов по каждому условию Количество голосов по каждому условию	Все	Обязательное
Сведения о субусловиях Перечень факторов, определяющих субусловие Значения MOS по субусловиям Количество голосов и дисперсия по каждому субусловию	Все	Необязательное
Воспроизведение аудиосигнала Ширина полосы аудиосигнала Каналы аудиосигнала (моно, стерео и т. д.) Уровень воспроизведения аудиосигнала Способ воспроизведения аудиосигнала, например через громкоговорители, наушники (моноуральный, диотический бинауральный и т. д.)	Речь, аудио, аудио + видео	Обязательное
Воспроизведение видеосигнала Разрешения видеоизображений (Примечания 1 и 2) Расстояние просмотра как функция высоты дисплея, например 3Н Тип и размеры устройства, например телевизор, планшет, телефон и т. д. Применение, например видеотелефония, видео по запросу, линейное ТВ и т. д.	Видео, аудио + видео	Обязательное
Язык	Пассивный Речь, аудио + видео	Обязательное
Число и пол говорящих субъектов	Пассивный Речь, аудио + видео	Обязательное
Тип видеоматериала, например спортивный сюжет или голова и туловище	Пассивный Видео, аудио + видео	Обязательное
Тип аудиоматериала, например классическая музыка, поп-музыка, саундтрек к фильму	Пассивный Речь, аудио + видео	Обязательное
ПРИМЕЧАНИЕ 1. – Должно быть отмечено использование чересстрочной развертки. ПРИМЕЧАНИЕ 2. – Если в эксперименте использовалось несколько разрешений изображения, необходимо сообщить, как именно отображались более низкие разрешения – штатным образом или путем масштабирования до наивысшего разрешения в эксперименте.		

13 Представление объективных значений MOS

При представлении значения MOS, полученного по модели объективной оценки МСЭ-Т, в общем случае достаточно сообщить, какая модель использовалась, а также какие были отклонения от стандартных (установленных по умолчанию) параметров. В случае нестандартных моделей необходимо привести информацию, указанную в строке "Методика" таблицы 1, для описания плана тестируемого эксперимента. Рекомендуется также привести сведения о типе тестируемого материала, использовавшегося в экспериментах для испытания модели объективной оценки и/или инструктирования по ее работе.

14 Условные обозначения

В [ITU-T P.800.1] введены общие и высокоуровневые условные обозначения, которые могут помочь в определении источника того или иного значения MOS. Эти обозначения в P.800.1 дают общий контекст, в котором было получено конкретное значение MOS, но не могут заменить подробное описание контекста, составленное в соответствии с настоящей Рекомендацией, которое следует предоставлять при любой возможности.

Библиография

МСЭ стандартизировал целый ряд методов субъективной оценки, предназначенных для различных целей. Ниже перечисляются некоторые наиболее широко применяемые из них.

[b-ITU-T G.729] Recommendation ITU-T G.729 (2012), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)*

Серия МСЭ-Т Р.800 включает в себя многочисленные Рекомендации, посвященные субъективной и объективной оценке качества речи. Особо упоминаются следующие из них.

[b-ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*

[b-ITU-T P.805] Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*

[b-ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*

[b-ITU-T P.862.1] Recommendation ITU-T P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO*

[b-ITU-T P.863] Recommendation ITU-T P.863 (2014), *Perceptual objective listening quality assessment*

Серия МСЭ-Т Р.900 включает в себя Рекомендации, посвященные оценке мультимедийных сигналов.

[b-ITU-T P.910] Рекомендация МСЭ-Т Р.910 (2008 год), *Методы субъективной оценки качества видеоизображения для мультимедийных приложений*

[b-ITU-T P.911] Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*

[b-ITU-T P.912] Recommendation ITU-T P.912 (2016), *Subjective video quality assessment methods for recognition tasks*

Сектором МСЭ-Р опубликованы также Рекомендации, посвященные субъективной оценке качества аудио- и видеосигналов.

[b-ITU-R BS.1116-1] Recommendation ITU-R BS.1116-1 (1997), *Methods for the subjective assessment of small impairments in audio systems including multichannel sounds systems*

[b-ITU-R BS.1534-1] Recommendation ITU-R BS.1534-1 (2003), *Method for the subjective assessment of intermediate quality levels of coding systems*

[b-ITU-R BT.500-13] Рекомендация МСЭ-Р BT.500-13 (2012 год), *Методика субъективной оценки качества телевизионных изображений*

[b-ITU-R BT.710-4] Recommendation ITU-R BT.710-4 (1998), *Subjective assessment methods for image quality in high-definition television*

ПРИМЕЧАНИЕ. – Не все стандартизированные методики, описываемые Сектором МСЭ-Р, предусматривают измерение средних экспертных оценок. Опубликованные Рекомендации МСЭ-Р содержат полный набор документации и справочных материалов по всем изложенным методикам. За более подробными описаниями и пояснениями указанных выше методик испытаний читателю предлагается обратиться к соответствующим отдельным Рекомендациям МСЭ-Р.

Субъективные методы испытаний и передовой опыт в этой области подробно рассматриваются в следующем справочнике.

[b-ITU-T handbook] *Практические процедуры для проведения субъективных испытаний* (2011 год)

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	Окружающая среда и ИКТ, изменение климата, электронные отходы, энергоэффективность; 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	Оконечное оборудование, субъективные и объективные методы оценки
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, сети последующих поколений, интернет вещей и "умные" города
Серия Z	Языки и общие аспекты программного обеспечения для систем электросвязи