

МСЭ-R

Сектор радиосвязи МСЭ

Рекомендация МСЭ-R ВТ.2095-1

(06/2017)

Субъективная оценка качества видео с использованием протокола просмотра экспертами

Серия ВТ

**Радиовещательная служба
(телевизионная)**



Предисловие

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

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

Политика в области прав интеллектуальной собственности (ПИС)

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

Серии Рекомендаций МСЭ-R

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

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

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

Электронная публикация
Женева, 2018 г.

© ITU 2018

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

РЕКОМЕНДАЦИЯ МСЭ-R ВТ.2095-1

Субъективная оценка качества видео с использованием протокола просмотра экспертами

(2016-2017)

Сфера применения

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

Ключевые слова

Телевидение, качество видео, субъективная оценка, просмотр экспертами

Ассамблея радиосвязи МСЭ,

учитывая,

- a)* что технологии кодирования источника для цифровых телевизионных приложений постоянно совершенствуются как по показателям эффективности, так и по зрительным характеристикам;
- b)* что непрерывное развитие технологий кодирования видео предполагает постоянно возрастающую потребность в методах анализа для оценки технических и зрительных характеристик;
- c)* что эффективность сжатия и зрительные характеристики, обеспечиваемые новыми технологиями кодирования видеоисточника, требуют новых и более эффективных методов оценки зрительных характеристик;
- d)* что методы оценки, определенные в действующих Рекомендациях МСЭ-R, требуют значительного времени и людских ресурсов и что зачастую в них не учитывается техническая эволюция дисплеев и развитие технических аспектов удовлетворенности конечного пользователя;
- e)* что новые подходы к получению протоколов просмотра экспертами продемонстрировали недавно более высокую эффективность и производительность по показателям времени и общих затрат по сравнению с методами, основанными на использовании не являющихся экспертами зрителей;
- f)* что, если результаты протокола просмотра экспертами не могут рассматриваться в качестве замены результатов, полученных на основании протокола официальной субъективной оценки, результаты протокола просмотра экспертами могут рассматриваться как ценное предварительное представление характеристик испытываемых систем;
- g)* что набирающая темпы технологическая эволюция в области дисплеев с плоским экраном коренным образом изменила условия, в которых обычно проводят просмотр эксперты;
- h)* что ИСО/МЭК при оценке новых технологий кодирования видеоисточника уже успешно используют новые протоколы на основе просмотра экспертами,

рекомендует,

1 чтобы при оценке новых цифровых технологий кодирования видеоисточника рассматривалось использование протокола просмотра экспертами, который описан в Приложении 1;

2 чтобы составление протокола просмотра экспертами осуществлялось с использованием профессиональных дисплеев с плоским экраном и в лабораторных условиях, которые описаны в Приложении 1.

ПРИМЕЧАНИЕ 1. – Приложение 2 (информационное) содержит результаты эксперимента по субъективной оценке с использованием протокола на основе просмотра экспертами, упомянутого в пункте *h)* раздела *учитывая*.

Приложение 1

Протокол просмотра экспертами для оценки качества видеоматериала

1 Организация лабораторных условий

1.1 Выбор дисплея и настройка

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

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

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

1.2 Расстояние просмотра

Расстояние просмотра, на котором располагаются эксперты, следует выбирать в соответствии с разрешающей способностью экрана и высотой активной части экрана, в соответствии с проектным расстоянием просмотра, которое описано в Рекомендации МСЭ-R ВТ.2022, или выбирать меньшее расстояние просмотра в соответствии с требованиями, определяемыми критическими условиями просмотра.

1.3 Условия просмотра

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

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

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

2 Зрители

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

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

Минимальное число разных зрителей должно составлять девять.

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

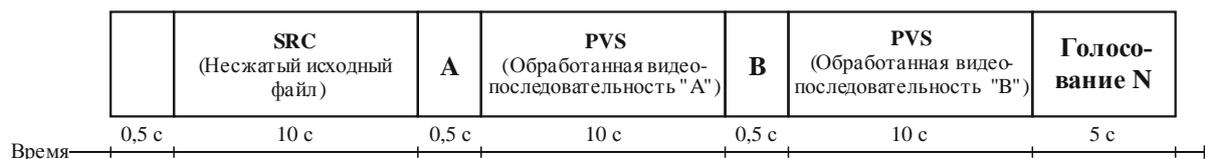
3 Базовая ячейка испытаний

Представляемый экспертам материал должен быть организован с созданием базовой ячейки испытаний (ВТС) для каждой пары оцениваемых условий кодирования (см. рисунок 1).

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

РИСУНОК 1

Временные интервалы базовой ячейки испытаний для протокола просмотра экспертами



ВТ.2095-01

ВТС должна быть организована следующим образом:

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

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

3.1 Оценочный лист и шкала оценок

Как показано на рисунке 1, представление видеофрагментов должно быть организовано таким образом, чтобы первой была показана неухудшенная эталонная последовательность (SRC), а далее – две ухудшенные видеопоследовательности (PVS). Порядок представления PVS должен произвольно меняться для каждой ВТС, и зрители не должны знать порядок представления.

РИСУНОК 2

Пример оценочного листа для состоящего из 24 ВТС сеанса просмотра экспертами

Сеанс 1

Голосование 1 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 2 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 3 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 4 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 5 А В <input type="checkbox"/> <input type="checkbox"/>
Голосование 6 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 7 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 8 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 9 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 10 А В <input type="checkbox"/> <input type="checkbox"/>
Голосование 11 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 12 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 13 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 14 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 15 А В <input type="checkbox"/> <input type="checkbox"/>
Голосование 16 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 17 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 18 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 19 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 20 А В <input type="checkbox"/> <input type="checkbox"/>
Голосование 21 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 22 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 23 А В <input type="checkbox"/> <input type="checkbox"/>	Голосование 24 А В <input type="checkbox"/> <input type="checkbox"/>	

Место
1 2 3

Участник

ВТ.2095-02

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

В таблице 1 показаны значения 11-балльной числовой шкалы.

ТАБЛИЦА 1

Значения 11-балльной числовой шкалы

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

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

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

Для каждой ВТС зрители заполняют ячейку, обозначенную буквой А (для оценки видеофрагмента, показанного первым), и ячейку, обозначенную буквой В (для оценки видеофрагмента, показанного вторым).

Представление исходного неухудшенного видеофрагмента упрощает оценку экспертами любых ухудшений.

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

3.2 План испытания и формирование сеанса

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

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

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

3.3 Подготовка участников

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

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

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

4 Сбор и обработка данных

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

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

Ко всем оценкам каждого участника следует применять функцию "корреляция" относительно показателя усредненного мнения (MOS); может быть установлено пороговое значение для определения каждого зрителя как "годный" и "негодный" (в Рекомендации МСЭ-Т Р.913 предлагается использовать пороговое значение "негодный", равное 0,75).

5 Условия использования результатов протокола просмотра экспертами

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

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

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

6 Ограничения использования результатов EVP

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

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

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

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

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

Приложение 2 (информационное)

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

В настоящем информационном приложении представлены сведения о результатах двух разных сеансов субъективной оценки кодированных видеофрагментов высокой четкости (ВЧ) и сверхвысокой четкости (СВЧ) для получения EVP, которые проводились в ходе 117-го собрания MPEG с применением положений Рекомендации МСЭ-R ВТ.2095, с тем чтобы быстро и надежно оценить два разных метода кодирования источника.

Ввиду присутствия большого числа экспертов, участвующих в 117-м собрании MPEG, количество оценщиков, участвовавших в двух сеансах для получения EVP, было значительно больше девяти – рекомендованного значения в Рекомендации МСЭ-R ВТ.2095; в тестовом сеансе оценки ВЧ для получения EVP приняли участие 30 экспертов, а в сеансе оценки СВЧ для получения EVP – 32 эксперта.

Широкое участие экспертов-оценщиков предоставило возможность проанализировать данные MOS с целью проверки уровня надежности, обеспечиваемого при использовании Рекомендации МСЭ-R ВТ.2095 для оценки кодированных видеофрагментов.

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

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

Как видно из рисунка 3 (эксперимент с видеоматериалом СВЧ) и рисунка 4 (эксперимент с видеоматериалом ВЧ), результаты оценки во всех четырех рассмотренных случаях весьма близки.

Приняв результаты, полученные от 18 зрителей, за своего рода "реальную ситуацию", можно построить графики, представленные на рисунках 3 и 4, на которые нанесены контрольные точки в соответствии с полученными значениями MOS от 18 зрителей (сплошная красная линия).

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

Анализируя результаты, представленные на рисунках 3 и 4, можно отметить:

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

В итоге, описанные здесь эксперименты для получения EVP показывают очень хорошую характеристику протокола EVP, подтверждая то, что заявлено в тексте Рекомендации МСЭ-R ВТ.2095, а именно что протокол EVP, хотя и не может рассматриваться в качестве полной замены официального субъективного эксперимента, он может рассматриваться в качестве стабильной процедуры оценки, обеспечивающей результаты, весьма близкие к тем, которые получены с участием гораздо большего числа зрителей при проведении официальной субъективной оценки.

РИСУНОК 3

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

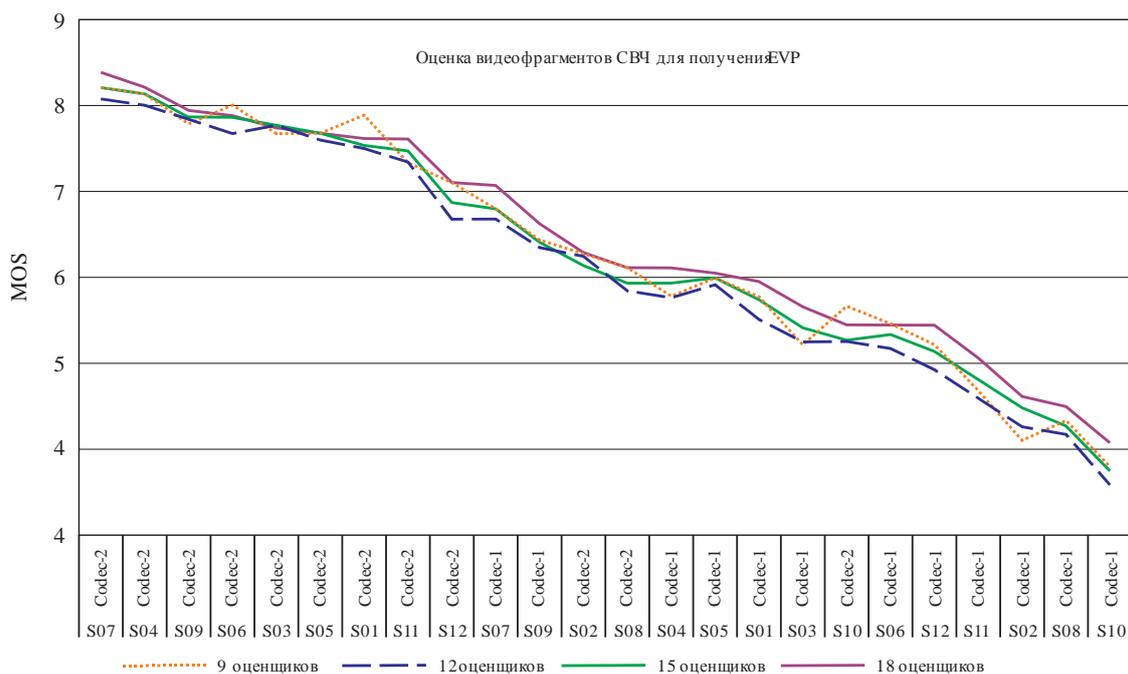
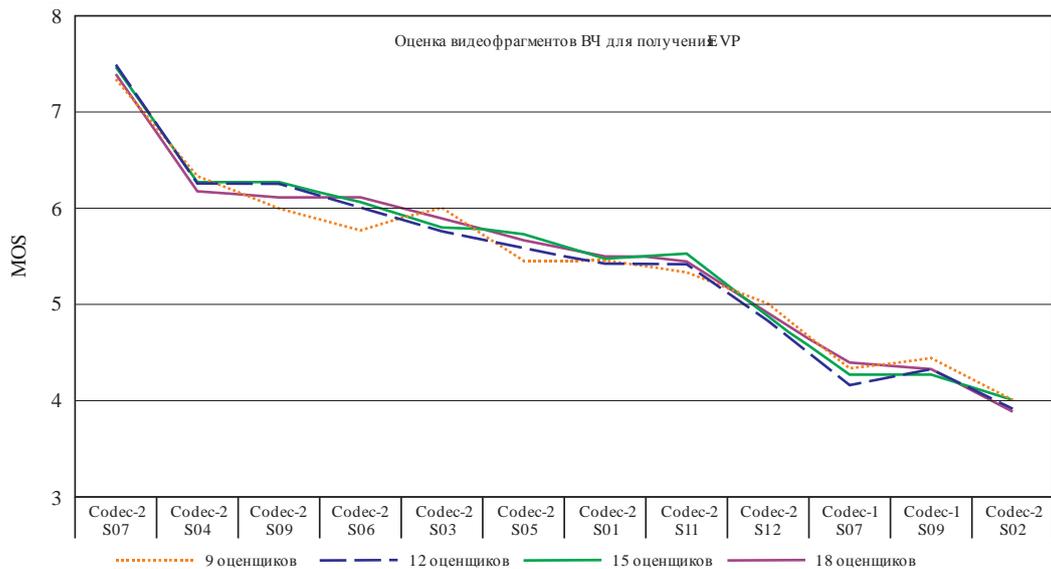


РИСУНОК 4

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



ВТ.2095-04