

Рекомендация МСЭ-R BS.1387-2

(05/2023)

Серия BS: Радиовещательная служба (звуковая)

**Методика объективных измерений
воспринимаемого качества звука**

Предисловие

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

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

Политика в области прав интеллектуальной собственности (ПИС)

Политика МСЭ-R в области ПИС излагается в общей патентной политике МСЭ-T/МСЭ-R/ИСО/МЭК, упоминаемой в 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/ru>, где также содержатся Руководящие принципы по выполнению общей патентной политики МСЭ-T/МСЭ-R/ИСО/МЭК и база данных патентной информации МСЭ-R.

Серии Рекомендаций МСЭ-R

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

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, 2024 г.

© ITU 2024

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 BS.1387-2

Методика объективных измерений воспринимаемого качества звука

(1998-2001-2023)

Сфера применения

В данной Рекомендации определена методика объективных измерений воспринимаемого качества звука.

Ключевые слова

Воспринимаемое качество звука, объективные измерения, кодирование с низкой битовой скоростью

Ассамблея радиосвязи МСЭ,

учитывая,

- a)* что принятые методы объективного измерения (например, отношения сигнал/шум и искажений) уже не подходят для измерения воспринимаемого качества звука в системах, в которых используются схемы кодирования с низкой битовой скоростью или аналоговая либо цифровая обработка сигналов;
- b)* что быстро внедряются схемы кодирования с низкой битовой скоростью;
- c)* что не все разработки, соответствующие спецификации или стандарту, гарантируют наилучшее качество в пределах таких спецификации или стандарта;
- d)* что формальные методы субъективной оценки не подходят для непрерывного контроля качества звука, например, в условиях эксплуатации;
- e)* что методы объективного измерения воспринимаемого качества звука могут фактически дополнять или заменять принятые методы объективного измерения во всех областях измерений;
- f)* что объективное измерение воспринимаемого качества звука может успешно дополнять методы субъективной оценки;
- g)* что для некоторых применений необходим метод, который можно реализовать в режиме реального времени,

рекомендует

- 1 для объективного измерения воспринимаемого качества звука в каждом из применений, перечисленных в Приложении 1, использовать метод, указанный в Приложении 2.

Предисловие

В настоящей Рекомендации определена методика объективных измерений воспринимаемого качества звука испытываемого устройства, например кодека с низкой битовой скоростью. Она состоит из двух приложений. В Приложении 1 приведен общий обзор методики для пользователей, и оно включает четыре прилагаемых документа. Прилагаемый документ 1 содержит описание применений и исследуемых сигналов. В Прилагаемом документе 2 перечислены выходные переменные модели и обсуждаются ограничения области применения и точности. В Прилагаемом документе 3 дано краткое описание модели, а в Прилагаемом документе 4 описаны принципы и общие характеристики объективных методов измерения воспринимаемого качества звука.

В Приложении 2 представлено подробное описание методики для разработчиков с использованием двух версий психоакустической модели, разработанной на этапе интеграции, в которой объединены шесть моделей. В Прилагаемом документе 1 к Приложению 2 описан процесс проверки методики объективных измерений. В Прилагаемом документе 2 к Приложению 2 представлен обзор всех баз данных, которые использовались при разработке и проверке этой методики.

СОДЕРЖАНИЕ

Стр.

Приложение 1 – Обзор	7
1 Введение.....	7
2 Области применения.....	7
3 Версии	8
4 Субъективный подход	8
5 Разрешение и точность	10
6 Требования и ограничения	10
Прилагаемый документ 1 к Приложению 1 – Применения.....	10
1 Общие положения	10
2 Основные применения	11
2.1 Оценка реализаций.....	11
2.2 Построение шкалы воспринимаемого качества	11
2.3 Мониторинг в онлайн-режиме	11
2.4 Состояние оборудования или соединения	11
2.5 Идентификация кодека	12
2.6 Разработка кодека.....	12
2.7 Планирование сети.....	12
2.8 Помощь в субъективной оценке	12
2.9 Краткое описание применений	13
3 Исследуемые сигналы.....	13
3.1 Выбор натуральных исследуемых сигналов.....	13
3.2 Длительность	14
4 Синхронизация	15
5 Вопросы авторских прав	15
Прилагаемый документ 2 к Приложению 1 – Выходные переменные	15
1 Введение.....	15
2 Выходные переменные модели.....	15
3 Базовое качество звука	15
4 Запас по кодированию	16
5 Требования пользователей	17

Прилагаемый документ 3 к Приложению 1 – Схема модели.....	17
1 Обработка звука	18
1.1 Пользовательские настройки	18
1.2 Психоакустическая модель.....	18
1.3 Когнитивная модель.....	19
Прилагаемый документ 4 к Приложению 1 – Принципы и характеристики объективных методов измерения воспринимаемого качества звука.....	19
1 Введение и предыстория	19
2 Общая структура объективных методов измерения воспринимаемого качества звука....	20
3 Психоакустические и когнитивные основы	20
3.1 Передаточная характеристика наружного и среднего уха	21
3.2 Шкалы воспринимаемых частот	21
3.3 Возбуждение	22
3.4 Обнаружение.....	23
3.5 Маскирование	23
3.6 Громкость и частичное маскирование.....	24
3.7 Резкость звука.....	24
3.8 Когнитивная обработка.....	24
4 Встроенные модели.....	25
4.1 DIX.....	25
4.2 NMR.....	26
4.3 OASE	26
4.4 Показатель воспринимаемого качества звука (PAQM)	27
4.5 Алгоритм PERCEVAL	27
4.6 Модель POM.....	28
4.7 Инструментальный метод.....	29
Приложение 2 – Описание модели	30
1 Основные принципы	30
1.1 Базовая версия	31
1.2 Расширенная версия.....	31
2 Модель периферического уха	32
2.1 Модель уха на основе БПФ	32
2.2 Модель уха на основе набора фильтров.....	43

3	Предварительная обработка шаблонов возбуждения	51
3.1	Адаптация уровня и шаблона.....	51
3.2	Модуляция	53
3.3	Громкость.....	53
3.4	Расчет сигнала ошибки	54
4	Расчет выходных переменных модели.....	54
4.1	Обзор	54
4.2	Разность показателей модуляции.....	54
4.3	Громкость шума	56
4.4	Ширина полосы	57
4.5	Отношение шума к маске	58
4.6	Доля кадров с помехами	59
4.7	Вероятность обнаружения.....	59
4.8	Гармоническая структура сигнала ошибки.....	61
5	Усреднение	62
5.1	Спектральное усреднение.....	62
5.2	Усреднение по времени	62
5.3	Усреднение по аудиоканалам.....	64
6	Оценка воспринимаемого базового качества звука	64
6.1	Искусственная нейронная сеть.....	64
6.2	Базовая версия	64
6.3	Расширенная версия.....	66
7	Соответствие реализаций	68
7.1	Общие положения	68
7.2	Выбор.....	68
7.3	Настройки для испытания на соответствие	68
7.4	Приемлемый интервал допуска	68
7.5	Исследуемые элементы.....	69
	Прилагаемый документ 1 к Приложению 2 – Процесс проверки.....	70
1	Общие сведения	70
2	Этап конкуренции	71
3	Этап сотрудничества.....	71

4	Проверка пригодности	72
4.1	Сравнение значений SDG и ODG	73
4.2	Корреляция.....	73
4.3	Абсолютная оценка ошибки (AES).....	75
4.4	Сравнение ODG с доверительным интервалом	77
4.5	Сравнение ODG с интервалом допуска.....	81
5	Выбор оптимальных версий модели	82
5.1	Критерии предварительного отбора, основанные на корреляции	82
5.2	Анализ количества выбросов	83
5.3	Анализ величины выбросов	84
6	Выводы.....	85
	Прилагаемый документ 2 к Приложению 2 – Описание справочных баз данных.....	85
1	Введение.....	85
2	Элементы, содержащиеся в базах данных	87
3	Условия экспериментов.....	87
3.1	MPEG90.....	88
3.2	MPEG91	88
3.3	ITU92DI.....	88
3.4	ITU92CO.....	88
3.5	ITU93	88
3.6	MPEG95.....	89
3.7	EIA95	89
3.8	DB2	89
3.9	DB3	89
3.10	CRC97.....	90
4	Элементы из DB2 и DB3 для каждого условия	90
4.1	DB2	90
4.2	DB3	93
	Справочные документы	96
	Библиография	97

Приложение 1

Обзор

1 Введение

Качество звука – один из ключевых факторов при проектировании цифровой системы радиовещания. Быстрое внедрение различных схем понижения битовой скорости привело к значительным усилиям по созданию и совершенствованию процедур субъективной оценки просто потому, что формальные испытания с прослушиванием оказались единственным подходящим методом оценки качества звука. Полученный опыт лег в основу Рекомендации МСЭ-R BS.1116, которая, в свою очередь, стала основой для большинства испытаний с прослушиванием этого типа.

Поскольку для субъективной оценки качества требуется много времени и средств, желательно разработать методику объективных измерений для оценки качества звука. На самом деле так и не было доказано, что традиционные методы объективных измерений, такие как измерение отношения сигнал/шум (S/N) или полного коэффициента гармонических искажений (THD), надежно связаны с воспринимаемым качеством звука. Проблемы становятся еще более очевидными, когда эти методы применяются к современным нелинейным и нестационарным кодекам.

За последнее десятилетие был внедрен ряд методов объективных перцептивных измерений воспринимаемого качества звука. Но ни один из этих методов не был тщательно проверен и, следовательно, не стал стандартизированным и не получил широкого распространения. В 1994 году МСЭ-R счел остро необходимым установить стандарт в этой области, и началась работа. Был объявлен открытый конкурс предложений и получены шесть возможных методов измерений: на основе индекса искажений (Disturbance Index – DIX), отношения шума к маске (Noise-to-Mask Ratio – NMR), показателя воспринимаемого качества звука (Perceptual Audio Quality Measure – PAQM), субъективной оценки (Perceptual Evaluation – PERCEVAL), воспринимаемого объективного показателя (Perceptual Objective Measure – POM), а также инструментальный метод. Эти методы описаны в Прилагаемом документе 4 к Приложению 1.

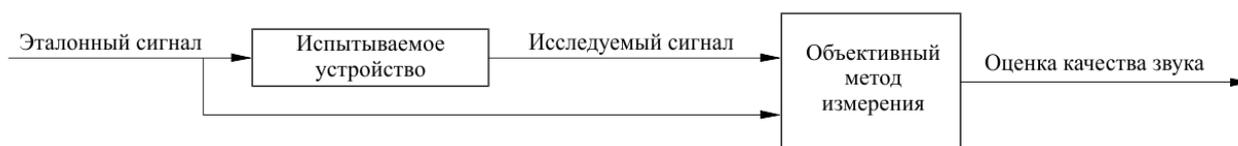
Методика измерения, описанная в настоящей Рекомендации, стала результатом процесса изучения эффективности каждого из шести вышеупомянутых методов, а также наиболее перспективных инструментов, выделенных и объединенных в единую методику. Рекомендуемая методика тщательно проверена на нескольких измерительных площадках. Доказано, что она дает надежную и полезную информацию для ряда применений. Однако следует иметь в виду, что методика объективных измерений, описанная в настоящей Рекомендации, в общем случае не заменяет организацию формальных испытаний с прослушиванием.

2 Области применения

Базовая концепция проведения объективных измерений рекомендуемым методом показана на рисунке 1.

РИСУНОК 1

Базовая концепция проведения объективных измерений



Методика измерения, описанная в настоящей Рекомендации, применима к большинству типов оборудования обработки аудиосигналов, как цифровых, так и аналоговых. Однако ожидается, что многие применения будут ориентированы на аудиокодеки.

Определены следующие восемь классов применений.

ТАБЛИЦА 1

Применения

	Применение	Краткое описание	Версия
1	Оценка реализаций оборудования	Процедура, характеризующая различные реализации оборудования для обработки звука, в основном аудиокодексов	Базовая/ расширенная
2	Построение шкалы воспринимаемого качества	Быстрая процедура, выполняемая перед вводом в эксплуатацию оборудования или схемы	Базовая
3	Непрерывный мониторинг	Непрерывный мониторинг передачи звука во время трансляции	Базовая
4	Состояние оборудования или соединения	Подробный анализ оборудования или схемы	Расширенная
5	Идентификация кодека	Процедура определения типа и реализации конкретного кодека	Расширенная
6	Разработка кодека	Процедура, максимально подробно характеризующая работу кодека	Базовая/ расширенная
7	Планирование сети	Процедура оптимизации стоимости и качества сети передачи данных при заданных ограничениях	Базовая/ расширенная
8	Помощь в субъективной оценке	Инструмент для отбора критически важных материалов для включения в программу испытаний с прослушиванием	Базовая/ расширенная

3 Версии

Чтобы добиться оптимального соответствия различным требованиям по стоимости и качеству, в настоящей Рекомендации предлагаются две версии методики объективного измерения. Базовая версия рассчитана на экономически эффективное внедрение для работы в режиме реального времени, тогда как расширенная версия ориентирована на достижение максимально возможной точности. В зависимости от реализации эта дополнительная точность повышает сложность методики примерно вчетверо по сравнению с базовой версией.

Некоторые рекомендации относительно того, какую версию применять в каждом случае, приведены в таблице 1.

4 Субъективный подход

Формальные субъективные испытания с прослушиванием, например основанные на Рекомендации МСЭ-R BS.1116, тщательно планируются таким образом, чтобы максимально приблизиться к надежной оценке качества звука. Однако нельзя ожидать, что результат субъективных испытаний с прослушиванием будет полностью отражать особенности фактического восприятия. Недостатки, присущие как субъективному, так и объективному подходу, иллюстрируются на рисунке 2.

Очевидно, что непосредственная проверка объективного метода невозможна. Вместо этого объективные методы измерения проверяются посредством субъективных испытаний с прослушиванием.

РИСУНОК 2
Способы проверки



BS.1387-02

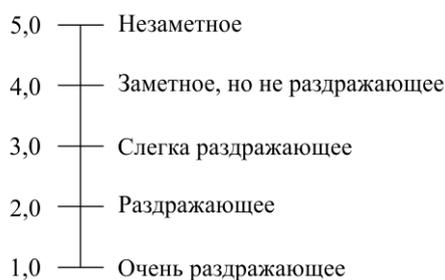
Методика объективных измерений, рассматриваемая в настоящей Рекомендации, ориентирована на применения, которые обычно оцениваются в рамках субъективного подхода с использованием Рекомендации МСЭ-R BS.1116. Основной принцип этой конкретной методики испытаний можно кратко описать следующим образом: слушатель может выбирать между тремя сигналами ("А", "В" и "С"). В качестве сигнала "А" всегда используется известный эталонный сигнал. Одновременно предлагаются скрытый эталонный сигнал и исследуемый сигнал, но им случайным образом, в зависимости от испытания, присваиваются имена "В" и "С").

Слушателю предлагается оценить ухудшение качества сигнала "В" по сравнению с сигналом "А" и ухудшение качества сигнала "С" по сравнению с сигналом "А" по непрерывной пятибалльной шкале. Один из сигналов – "В" или "С" – должен быть неотличим от сигнала "А"; другой может демонстрировать ухудшение качества. Любые воспринимаемые различия между эталоном и другими сигналами должны трактоваться как ухудшение. Обычно используется только один атрибут – "базовое качество звука". Он определяется как глобальный атрибут, который включает в себя все обнаруженные различия между эталонным и исследуемым сигналами.

Шкала оценки считается непрерывной с опорными точками, соответствующими принятой МСЭ-R пятибалльной шкале ухудшения качества, которая представлена в Рекомендации МСЭ-R BS.1284, как указано ниже.

РИСУНОК 3

Пятибалльная шкала ухудшения качества звука МСЭ-R



BS.1387-03

Анализ результатов субъективных испытаний с прослушиванием обычно основан на шкале субъективной оценки разницы (Subjective Difference Grade – SDG), определяемой следующим образом:

$$SDG = Grade_{Signal\ Under\ Test} - Grade_{Reference\ Signal}.$$

В идеале значения SDG должны находиться в диапазоне от 0 до –4, где 0 соответствует незаметным ухудшениям, а –4 – ухудшениям, которые оцениваются как очень раздражающие.

5 Разрешение и точность

Выходной переменной объективного метода измерения является шкала объективной оценки разницы (Objective Difference Grade – ODG), которая соответствует SDG при субъективном подходе. Разрешение ODG ограничено одним десятичным знаком. Однако следует соблюдать осторожность, не ожидая, что в общем случае разница между любой парой ODG в одну десятую балла будет ощутимой. То же замечание справедливо и при рассмотрении результатов субъективных испытаний с прослушиванием.

Единого показателя, полностью описывающего точность объективного метода измерения, не существует. Вместо этого необходимо учитывать ряд разных показателей качества. Одним из них служит корреляция между SDG и ODG. Важно понимать, что нет никакой гарантии, что эта корреляция превысит некоторое заданное значение. Эффективность метода измерения скорее всего будет зависеть, например, от типа и уровня внесенного ухудшения качества.

Еще одним показателем качества, представляющим интерес, является количество выбросов. Выброс определяется как измеренное значение, не соответствующее заданной схеме допусков. Согласно требованиям пользователей метод измерения должен обеспечивать максимально возможную точность в верхней части шкалы оценок (то есть для высококачественного звука). Следовательно, в средней и нижней частях шкалы оценок полученная точность может быть ниже.

Хотя корреляция обычно дает хорошую оценку точности объективного метода измерения, важно иметь в виду, что даже за относительно высоким показателем корреляции может скрываться неприемлемое качество метода измерения (с точки зрения выбросов).

Третьим показателем качества, который использовался в процессе проверки, служит абсолютная оценка ошибки (Absolute Error Score – AES), отражающая среднее значение отношения между размером доверительного интервала SDG и разницей между SDG и ODG.

Более подробная информация об ожидаемых характеристиках метода измерения, а также о характеристиках, полученных в процессе проверки, приведена в Прилагаемом документе 1 к Приложению 2.

6 Требования и ограничения

Сигнал от испытываемого устройства и эталонный сигнал должны быть синхронизированы по времени с точностью до 24 выборок в течение всего интервала измерения. Механизм синхронизации в настоящей Рекомендации не рассматривается и, как ожидается, будет различным в зависимости от реализации.

Прилагаемый документ 1 к Приложению 1

Применения

1 Общие положения

В этом Прилагаемом документе представлены определения и конкретные требования для основных применений рекомендуемой методики объективного измерения воспринимаемого качества звука.

Для некоторых применений требуется реализация объективного метода измерения в режиме реального времени, в то время как для других достаточно измерения не в реальном времени. Для реализаций реального времени рекомендуется, чтобы максимальная задержка в измерительном оборудовании не превышала 200 мс, и недопустима задержка более 1 с.

Кроме того, необходимо различать измерения в онлайнном и офлайнном режимах. При офлайн-измерениях сохраняется возможность полного доступа к оборудованию или соединению, а при онлайнных выполняется программа измерений, прерывать которую нельзя.

2 Основные применения

2.1 Оценка реализаций

Радиовещательным компаниям, сетевым операторам и другим организациям необходимо оценивать различные реализации оборудования, в частности аудиокодеков, при выборе такого оборудования для покупки или при проведении приемочных испытаний.

В такого рода применениях требуется высокая точность, особенно для оценки небольших ухудшений качества звука и правильной квалификации различных реализаций. Что касается выходных переменных, то пользователям достаточно простого результата, такого как ODG, но разработчики аудиокодеков могут провести более тщательный анализ, используя подходящий набор выходных переменных модели (Model Output Variables – MOV).

Можно использовать обе версии модели, но рекомендуется расширенная версия

2.2 Построение шкалы воспринимаемого качества

Это быстрая процедура, выполняемая перед вводом в эксплуатацию оборудования или схемы. Целью является проверка функциональных возможностей и качества. Измерительным оборудованием занимается эксплуатационный персонал. Могут присутствовать любые виды ухудшения качества звука.

Требуется измерение в режиме реального времени. Могут использоваться исследуемые сигналы или заранее определенные звуковые сигналы. ODG должны правильно отображаться и выдаваться не реже двух раз в секунду, или, если используется специальный исследуемый сигнал, непосредственно после окончания этого исследуемого сигнала.

Достаточно использовать базовую версию.

2.3 Мониторинг в онлайнном режиме

Это непрерывный процесс, происходящий во время непрерывной передачи звука. Программа не должна прерываться процедурой измерения. Следовательно, для измерения должен использоваться сигнал самой программы или заранее определенный аудиофрагмент. Последний может быть сигналом станции или музыкальной заставкой. Измерительным оборудованием занимается эксплуатационный персонал.

Требуется измерение в режиме реального времени. ODG должны отображаться надлежащим образом и выдаваться не реже двух раз в секунду или непосредственно после окончания заданного сигнала. Отображение MOV нежелательно.

Достаточно использовать базовую версию.

2.4 Состояние оборудования или соединения

Требуется периодическая тщательная проверка качества, гарантирующая функционирование соединений и звукового оборудования. В отличие от онлайнного мониторинга или построения шкалы воспринимаемого качества для этого применения требуется проверка нескольких технических параметров.

Измерительная система должна предоставлять подробную информацию о влиянии состояния оборудования или соединения на воспринимаемое качество звука, отображая в дополнение к ODG полный набор MOV. Измерение в режиме реального времени не требуется.

Рекомендуется использовать расширенную версию.

2.5 Идентификация кодека

Для того чтобы идентифицировать кодеки (разные алгоритмы или разные реализации одного и того же алгоритма), измерительная система должна быть способна хранить, извлекать и сравнивать наборы характеристик. Сходство между такими наборами характеристик (шаблонами) можно рассматривать как меру сходства различных реализаций кодеков. Такая процедура используется для определения типа и реализации конкретного кодека.

Измерительная система должна записывать как можно больше информации о шаблонах. Рассмотрение только ODG может не дать достаточной информации.

Достаточно использовать базовую версию, хотя измерения в режиме реального времени не требуются.

ПРИМЕЧАНИЕ. – Существует ограниченный опыт применения рекомендуемого метода. Более того, до сих пор не определена единая мера сходства между шаблонами.

2.6 Разработка кодека

Метод измерения для этой цели должен давать как можно более точную и подробную характеристику испытываемого кодека, особенно при малых искажениях.

Для непрерывного мониторинга требуется обработка в режиме реального времени, которая может не поддерживаться расширенной версией. Однако при небольших искажениях и для получения подробной информации потребуется расширенная версия. Система измерения должна быть способна отображать выходные данные с той же скоростью, с какой они рассчитываются. Желателен прямой доступ к истории выходных данных за период 4 с.

Рекомендуется использовать расширенную версию. Однако для измерений в режиме реального времени достаточно базовой версии. Требуется анализ как в режиме реального времени, так и не в режиме реального времени, а также покадровый анализ. Любое значительное искажение должно быть указано, например, с помощью индикатора пиковых значений. Желателен доступ к полному набору MOV.

2.7 Планирование сети

Для планирования сети требуется оценка ожидаемого качества звука на разных этапах процесса планирования. Для изучения различных конфигураций в целях оптимизации качества звука можно использовать программное моделирование компонентов сети, позволяющее комбинировать различные этапы обработки звука. На более позднем этапе можно протестировать фактические компоненты обработки звука в выбранной конфигурации.

Планирование сети выполняется инженерами-системотехниками, которые должны получить подробную информацию о влиянии характеристик сети на качество звука. Оценка возможных конфигураций сети должна основываться на подходящем наборе MOV в зависимости от конкретного применения сети. Таким образом, отображения только ODG недостаточно. Измерения в режиме реального времени для оценки качества в этом применении не требуются.

Можно использовать обе версии модели, но рекомендуется расширенная версия.

2.8 Помощь в субъективной оценке

Метод объективного измерения представляет собой инструмент для отбора критически важного аудиоматериала, который будет использоваться в субъективных испытаниях с прослушиванием. Для категоризации такого материала можно использовать весь набор MOV.

Требуется максимально возможная точность, и рекомендуется использовать расширенную версию. Однако желательно измерение в режиме реального времени, чтобы сократить время, необходимое для отбора критически важного материала.

2.9 Краткое описание применений

Требования, предъявляемые к методике измерения для основных применений, обобщены в таблице 2.

ТАБЛИЦА 2
Требования к методике измерения

	Применение	Категория	Режим реального времени	Мин., ROV ⁽¹⁾ (Гц)	Онлайн/офлайн	Версия модели
1	Оценка реализаций	Диагностика	Нет	–	Офлайн	Обе
2	Построение шкалы воспринимаемого качества	Эксплуатация	Да/нет	2	Офлайн	Базовая
3	Мониторинг в онлайн-режиме	Эксплуатация	Да	2	Онлайн	Базовая
4	Состояние оборудования или соединения	Диагностика	Да/нет	–	Онлайн/офлайн	Расширенная
5	Идентификация кодека	Диагностика	Нет	–	Офлайн	Обе
6	Разработка кодека	Разработка	Да/нет	–	Офлайн	Обе
7	Планирование сети	Разработка	Да/нет	–	Офлайн	Обе
8	Помощь в субъективной оценке	Разработка	Да/нет	–	Офлайн	Расширенная

⁽¹⁾ Скорость вывода значений (в секунду).

3 Исследуемые сигналы

Исследуемые сигналы можно разделить на две группы – натуральные и синтетические. Представленный здесь список натуральных исследуемых сигналов состоит из важнейших звуковых последовательностей, уже использовавшихся в испытаниях с прослушиванием, проводившихся как МСЭ-R, так и другими организациями для оценки качества звука. Сигналы должны быть доступны как в месте передачи, так и в месте измерения. Таким образом, требуется измерительное устройство с возможностью запоминания.

Синтетические сигналы определяются математически и могут изменяться контролируемым образом. Такие сигналы можно генерировать в местах передачи и измерения. Дополнительная память в измерительном устройстве не требуется. Дать субъективную оценку таким сигналам ввиду их характера сложно, если вообще возможно. Поэтому метод измерения качества этих сигналов не проверялся на основе субъективных результатов.

3.1 Выбор натуральных исследуемых сигналов

В таблице 3 представлен список подмножества исследуемых сигналов, которые использовались в процедуре проверки пригодности, приведшей к созданию настоящей Рекомендации. Также указан тип артефактов, которые обычно обнаруживаются в этих сигналах из-за низкой битовой скорости кодирования.

ТАБЛИЦА 3

Подмножество исследуемых сигналов

№ п/п	Сигнал	Имя файла	Примечания
1	Кастаньеты	cas	(1)
2	Кларнет	cla	(2)
3	Клаве	clv	(1)
4	Флейта	flu	(2)
5	Колокольчики	glo	(1), (2), (5)
6	Клавесин	hrp	(1), (2), (4)
7	Литавры	ket	(1)
8	Маримба	mar	(1)
9	Фортепьяно, Шуберт	pia	(2)
10	Камертон-дудка	pip	(4)
11	Рай Кудер	rys	(2), (4)
12	Саксофон	sax	(2)
13	Вольтка	sb1	(2), (4), (5)
14	Речь женская, англ. язык	sfe	(3)
15	Речь мужская, англ. язык	sme	(3)
16	Речь мужская, нем. язык	smg	(3)
17	Малые барабаны	sna	(1)
18	Сопрано, Моцарт	sop	(4)
19	Бубен	tam	(1)
20	Труба	tpt	(2)
21	Треугольник	tri	(1), (2), (5)
22	Туба	tub	(2)
23	Сюзанна Вега	veg	(3), (4)
24	Ксилофон	xyl	(1), (2)

- (1) Переходные процессы: чувствительность к прерыванию, размытие шума во временной области.
- (2) Тональный состав: чувствительность к шуму, резкость.
- (3) Естественная речь (критическое сочетание тональных фрагментов и фронтов): чувствительность к искажениям, размытие фронтов.
- (4) Сложный звук: нагружает испытываемое устройство.
- (5) Широкая полоса пропускания: нагрузка на испытываемое устройство, потеря высоких частот, программно модулированный высокочастотный шум.

3.2 Длительность

Длительность натурального исследуемого сигнала должна быть примерно такой же, как если бы он использовался в испытании с прослушиванием. Обычно длительность составляет от 10 до 20 с. Весьма вероятно, что критически важная часть исследуемого сигнала, демонстрирующая большинство артефактов, составляет лишь его малую долю.

Длительность синтетических исследуемых сигналов должна быть достаточно большой, чтобы создать нагрузку на испытываемый кодек, который может содержать буфер для кодированного аудиосигнала. С учетом длины буфера и постоянных времени, характерных для метода измерения, длительность каждого отдельного исследуемого элемента в составе последовательности должна превышать 500 мс.

Длительность может быть ограничена таким коротким значением, поскольку не ожидается, что эти сигналы будут использоваться в субъективных испытаниях с прослушиванием.

4 Синхронизация

Исследуемый сигнал и эталонный сигнал для процедуры измерения должны быть синхронизированы во времени. Это справедливо как для натуральных, так и для синтетических исследуемых сигналов.

5 Вопросы авторских прав

Исследуемые сигналы, приведенные в таблице 3, могут использоваться без учета авторских прав только для целей измерения в сочетании с методом объективных измерений, описанным в Приложении 2 к настоящей Рекомендации.

ПРИМЕЧАНИЕ. – Для всех последовательностей, которые взяты главным образом у Европейского радиовещательного союза (диск SQAM EPC), должно быть получено разрешение на использование авторских прав.

Прилагаемый документ 2 к Приложению 1

Выходные переменные

1 Введение

Методика объективного измерения, описанная в настоящей Рекомендации, позволяет измерить качество звука и получить значение, соответствующее воспринимаемому качеству звука. Методика измерения моделирует фундаментальные свойства слуховой системы. Несколько промежуточных этапов моделируют физиологические и психоакустические эффекты.

Эти промежуточные выходные данные можно использовать для характеристики артефактов. Они называются выходными переменными модели (Model Output Variables – MOV). На заключительном этапе измерения модель объединяет значения MOV для получения единого выходного значения, которое непосредственно соответствует ожидаемому результату субъективной оценки качества.

2 Выходные переменные модели

Таблица 4 содержит описание MOV, используемых для прогнозирования объективных оценок разницы. Индексы_A выводятся из части модели, основанной на наборе фильтров, а индексы_B – из части модели, основанной на БПФ. Объективные оценки разницы можно предсказать либо только на основе БПФ (базовая версия), либо на основе комбинации БПФ и набора фильтров (расширенная версия). Усреднение всегда выполняется по времени.

3 Базовое качество звука

Наиболее известным параметром субъективных испытаний с прослушиванием является базовое качество звука (Basic Audio Quality – BAQ). BAQ измеряется как субъективная оценка разницы (SDG), которая рассчитывается как оценка, присвоенная исследуемому сигналу в ходе субъективного испытания¹, за вычетом оценки, присвоенной эталонному сигналу. SDG обычно имеет отрицательное

¹ См. Рекомендацию МСЭ-R BS.1116.

значение. Соответствующий выходной параметр модели называется объективной оценкой разницы (ODG). Сопоставление MOV с ODG основано на большом количестве надежных исследуемых элементов, см. Прилагаемый документ 2 к Приложению 2.

ТАБЛИЦА 4

Описание выходных переменных модели

Выходные переменные модели	Описание
<i>WinModDiff_B</i>	<i>Скользящее среднее разности показателей модуляции (огibaющих) эталонного и исследуемого сигналов</i>
<i>AvgModDiff1_B</i>	<i>Усредненная разность показателей модуляции</i>
<i>AvgModDiff2_B</i>	<i>Усредненная разность показателей модуляции с упором на внесенные модуляции и изменения модуляций, когда эталонный сигнал содержит мало модуляций или вообще не содержит их</i>
<i>RmsModDiff_A</i>	<i>Среднеквадратичное значение разности показателей модуляции</i>
<i>RmsMissingComponents_A</i>	<i>Среднеквадратичное значение громкости шума из-за недостающих частотных составляющих (используется в переменной <i>RmsNoiseLoudAsym_A</i>)</i>
<i>RmsNoiseLoud_B</i>	<i>Среднеквадратичное значение усредненной громкости шума с упором на вносимые компоненты</i>
<i>RmsNoiseLoudAsym_A</i>	<i><math>RmsNoiseLoud_A + 0,5RmsMissingComponents_{A}}</math></i>
<i>AvgLinDist_A</i>	<i>Показатель средних линейных искажений</i>
<i>BandwidthRef_B</i>	<i>Ширина полосы эталонного сигнала</i>
<i>BandwidthTest_B</i>	<i>Ширина полосы выходного сигнала испытываемого устройства</i>
<i>TotNMR_B</i>	<i>Логарифм усредненного отношения общего шума к маске</i>
<i>RelDistFrames_B</i>	<i>Относительная доля кадров, в которых хотя бы одна полоса частот содержит значительную шумовую составляющую</i>
<i>AvgSegmNMR_B</i>	<i>Сегментно усредненный логарифм отношения шума к маске</i>
<i>MFPD_B</i>	<i>Максимум вероятности обнаружения после фильтрации нижних частот</i>
<i>ADB_B</i>	<i>Средний искаженный блок (= кадр), взятый как логарифм отношения общего искажения к общему количеству кадров со значительными искажениями</i>
<i>EHS_B</i>	<i>Гармоническая структура ошибок по времени</i>

ODG – это объективно измеряемый параметр, соответствующий субъективно воспринимаемому качеству. Поскольку задачей слушателя при прослушивании исследуемого элемента является оценка BAQ испытываемого устройства, ODG также является показателем BAQ.

4 Запас по кодированию

Еще один параметр, который в будущем может оказаться ценным, – это запас по кодированию (Coding Margin – CM) – способ описания неслышимых артефактов. Субъективный запас по кодированию (Subjective Coding Margin – SCM) можно оценить путем усиления артефактов до тех пор, пока они не станут слышимыми для испытываемого. SCM описывает запас до порога слышимости артефактов.

Чтобы найти порог, артефакты необходимо усилить или ослабить во время прослушивания. Подходящим методом может служить разностный метод. Разностный сигнал синхронных по времени исходного и кодированного сигналов усиливается и добавляется к исходному сигналу. Определение порога слышимости лучше всего проводить методом принудительного выбора. SCM получается путем

усреднения пороговых значений степени усиления или ослабления, полученных от испытуемых. Отрицательные значения SM соответствуют слышимым артефактам, а положительные – неслышимым. В отличие от BAQ , запас по кодированию служит мерой того, когда (при каком уровне громкости) артефакты становятся слышимыми, а не того, насколько они раздражают. Процессы определения и проверки метода измерения SCM описаны в [Feiten, 1997].

Значение объективного запаса по кодированию (Objective Coding Margin – OCM) также можно получить из MOV . В настоящее время оценка субъективного запаса по кодированию проведена лишь для нескольких исследуемых элементов. Преобразование из модели, представленной в данной Рекомендации, в OCM еще не исследовалось.

5 Требования пользователей

Требования пользователей к выходным переменным метода измерения зависят от применения. Для некоторых применений, например под номерами 2 и 3 (см. Прилагаемый документ 1 к Приложению 1), измерение является частью рабочей процедуры. В этих случаях очень важно, чтобы результаты метода были легко читаемы и легко интерпретировались людьми, не обладающими глубокими знаниями о технике измерения. Проще всего этого достичь, если метод дает **только одно значение**, соответствующее воспринимаемому качеству звука.

То же может относиться и к другим применениям, например под номерами 1 и 4. Однако для них, а также для применений под номерами 5–8 могут быть полезны более сложные выходные переменные, ориентированные на пользователей с более глубокими знаниями в области механизмов измерения.

Прилагаемый документ 3 к Приложению 1

Схема модели

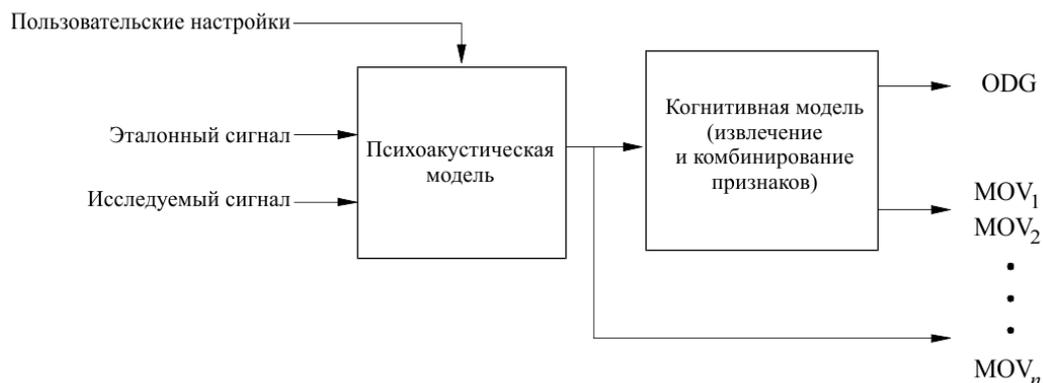
В соответствии с Рекомендацией МСЭ-R BS.1116 SDG для исследуемых звуковых элементов получают в ходе прослушивания, а за субъективное качество сигнала принимается усредненное значение SDG по нескольким слушателям. Элемент может содержать различные искажения звука, поэтому изменения качества интегрируются по времени. Следовательно, для прогнозирования SDG на основе физических измерений требуется точная модель периферической слуховой системы, а также когнитивных аспектов оценки качества звука.

Рекомендуемая модель для объективного измерения создает ряд выходных переменных модели (MOV) на основе сравнения эталонного сигнала с исследуемым. Эти MOV сопоставляются с ODG с использованием метода оптимизации, который минимизирует квадрат разности распределения ODG и соответствующего распределения средних значений SDG для достаточно большого набора данных.

Описаны два варианта модели: версия на основе ДПФ, которую можно использовать для мониторинга в режиме реального времени, и версия, основанная на наборе фильтров и ДПФ, которая, как ожидалось, даст более точные результаты. Версия на основе ДПФ считается базовой, а комбинированная версия – расширенной.

Общая структура базовой и расширенной версий показана на рисунке 4.

РИСУНОК 4
Этапы обработки, реализованные в модели



BS.1387-04

1 Обработка звука

Как и при субъективных испытаниях с прослушиванием, качество исследуемого сигнала оценивается по отношению к эталонному сигналу. Эталонный и исследуемый сигналы (монофонические или стереосигналы) преобразуются в психоакустическое представление. Эти представления сравниваются для получения ODG. Данные операции выполняются в ходе этапов обработки, показанных на рисунке 4.

1.1 Пользовательские настройки

Для методики измерения в качестве параметра требуется предполагаемый уровень громкости прослушивания. Поэтому пользователь должен указать уровень звукового давления в дБ SPL, создаваемого полномасштабной синусоидальной волной с частотой 1019,5 Гц. Если точный уровень громкости прослушивания неизвестен, рекомендуется принять уровень громкости прослушивания равным 92 дБ SPL.

1.2 Психоакустическая модель

Психоакустическая модель преобразует последовательные кадры сигнала во временной области в представление на основе движения базилярной мембраны. Этот процесс начинается с использования как ДПФ, так и набора фильтров. ДПФ переводит данные в частотную область, а результат преобразуется из частотной шкалы в тональную – психоакустический эквивалент частоты. В части модели, связанной с набором фильтров, превращение частоты в высоту тона непосредственно учитывается шириной полосы пропускания и расстоянием между полосовыми фильтрами.

Для одновременного маскирования используются две разные концепции. Некоторые значения MOV рассчитываются с использованием *концепции порога маскирования*, тогда как другие основаны на *сравнении внутренних представлений*. Первая концепция непосредственно рассчитывает порог маскирования с использованием функций психофизического маскирования. Выходные переменные модели основаны на удаленности сигнала физической ошибки до этого порога маскирования. При сравнении внутренних представлений энергия исследуемого сигнала (Signal Under Test – SUT) и эталонного сигнала распределяется по соседним областям основного тона для получения шаблонов возбуждения. Выходные переменные модели основаны на сравнении этих шаблонов возбуждения. Неодновременное маскирование реализуется путем размывания представлений сигналов во времени.

Абсолютный порог моделируется отчасти путем применения частотно-зависимой весовой функции и отчасти путем добавления к шаблонам возбуждения частотно-зависимого смещения. Этот порог представляет собой аппроксимацию минимального слышимого давления [ИСО 389-7, Акустика. Опорный нуль для калибровки аудиометрической аппаратуры. Часть 7: Опорный порог слышимости при прослушивании в условиях свободного и диффузного звуковых полей, 1996 год].

Основными результатами психоакустической модели являются возбуждение и порог маскирования как функция времени и частоты. Результаты модели на нескольких уровнях доступны для дальнейшей обработки.

1.3 Когнитивная модель

Когнитивная модель собирает воедино информацию из последовательности кадров, созданных психоакустической моделью. Наиболее важными источниками информации для проведения качественных измерений являются разность между эталонным сигналом и исследуемым сигналом как в частотной, так и в тональной области. В частотной области измеряются спектральные полосы обоих сигналов, а также гармоническая структура ошибки. В тональной области показатели ошибки определяются как по модуляции огибающей возбуждения, так и по величине возбуждения.

Рассчитанные характеристики взвешиваются так, что их комбинация приводит к получению ODG, достаточно близкой к SDG конкретного исследуемого ухудшения качества звука. В базовой версии для создания ODG используются 11 параметров, а в расширенной версии 5. Оптимизация проводилась с использованием алгоритма обратного распространения для обучения нейронной сети (см. пункт 6 Приложения 2). Данные для обучения состояли из полных баз данных 1 и 2 и части базы данных 3. Для обобщения данных испытаний использовались оставшаяся часть базы данных 3 и полный набор данных CRC97 (см. Прилагаемый документ 2 к Приложению 2).

Прилагаемый документ 4 к Приложению 1

Принципы и характеристики объективных методов измерения воспринимаемого качества звука

1 Введение и предыстория

Цифровая передача и хранение аудиосигналов все чаще основываются на алгоритмах сжатия данных, адаптированных к свойствам слуховой системы человека, и, в частности, полагаются на маскирующие эффекты. Такие алгоритмы направлены главным образом не на минимизацию искажений, а на обработку этих искажений таким образом, чтобы они воспринимались как можно слабее. Качество этих кодеков восприятия уже нельзя оценивать обычными методами, которые, как правило, определяют общую величину искажения. В качестве примера для иллюстрации этих ограничений часто приводится так называемое чудо 13 дБ: наложенный шум со спектральной структурой, адаптированной к звуковому сигналу, практически не слышен, даже когда результирующее невзвешенное соотношение сигнал/шум снижается до 13 дБ.

По этой причине для оценки качества звука перцептивных кодеков требуются испытания с прослушиванием. Для достижения достаточной надежности и повторяемости испытаний с прослушиванием требуются большие затраты времени и труда.

Решить эти проблемы помогают схемы объективных измерений, учитывающие свойства слуховой системы человека. Эта идея была впервые опубликована в [Schroeder *et al.*, 1979]. В этой работе, посвященной в основном кодированию речи, описана схема измерения громкости шума (NL).

Воспринимаемая громкость сигнала шума речевого кодека, выражающая разницу между входным и выходным сигналами, оценивалась для каждого временного кадра длительностью примерно 20 мс. Если сигнал шума полностью замаскирован, то воспринимаемая громкость считается равной нулю. Частичное маскирование понижает громкость незамаскированного шумового сигнала. Используемый порог маскирования оптимизирован для тонально-маскирующего шума, и для каждого кадра рассчитывается результирующее ухудшение качества речи. Общее качество речевого фрагмента не вычисляется.

В 1985 году Карьялайнен опубликовал схему измерения слышимой спектральной разницы (ASD) [Karjalainen, 1985]. Начав с нескольких идей Шредера, Атала и Холла, он заменил анализ на основе кадров набором фильтров с перекрывающимися фильтрами, изменил способ включения абсолютного порога и добавил модель временного маскирования. Оба входных сигнала схемы измерения обрабатываются одинаково, так что создается своего рода внутреннее представление. Эти внутренние представления сравниваются друг с другом, чтобы можно было объяснить воспринимаемые различия между входным и выходным сигналом схемы кодирования речи. Общее качество речевого фрагмента не вычисляется. Разрешение ASD по времени лучше адаптировано к свойствам слуховой системы человека, но приводит к усложнению алгоритма.

В 1987 году Бранденбург опубликовал схему измерения отношения шума к маске (NMR) [Brandenburg, 1987] для использования в качестве инструмента разработки схем кодирования звука. Эта схема проще схемы NL благодаря тому, что расчет распределения по диапазонам восприятия производится с использованием функции распределения на основе кривой наихудшего случая. Используемый порог маскирования оптимизирован для тонально-маскирующего шума. Были добавлены простая схема моделирования звука после маскирования и несколько способов оценки воспринимаемого качества более длительных звуковых фрагментов. Эта схема стала первой моделью, реализованной на аппаратуре реального времени.

В 1989 году Мур и Гласберг [Moore, 1989] представили модель восприятия, но не предложили способа оценки воспринимаемого качества искаженных аудиосигналов.

2 Общая структура объективных методов измерения воспринимаемого качества звука

Все схемы перцептивного измерения работают с двумя входными сигналами: эталонным (REF) и исследуемым (SUT). В ситуациях, когда эталонный сигнал не может передаваться в измерительное оборудование, но хорошо известен, он может быть внутренним эталонным сигналом, который хранится в самом измерительном оборудовании. Очень важно, чтобы входные сигналы были синхронизированы по времени.

Включение в схемы измерений психоакустики может осуществляться двумя способами. Первый очень похож на структуру схем кодирования звука: эталонный сигнал используется для расчета оценки фактического порога маскирования (см. ниже). Разница между исследуемым и эталонным сигналами сравнивается с этим порогом маскирования. Этот метод называется "концепция порога маскирования" и используется в измерениях громкости шума и NMR. Разницу между входными сигналами можно рассчитать либо во временной области, либо как разность между кратковременными энергетическими спектрами. В последнем случае обеспечивается лучшая устойчивость к ошибкам синхронизации, но снижается разрешение по времени. Разница во временной области обычно слишком чувствительна к фазовым искажениям и поэтому больше не используется.

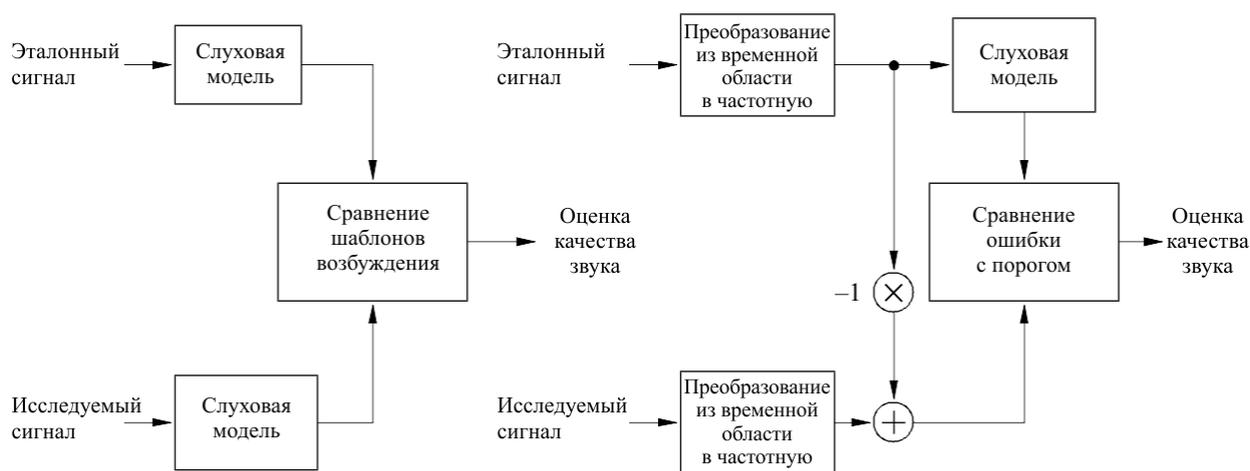
Второй подход ближе к физиологическим процессам в слуховой системе человека: рассчитывается так называемое внутреннее представление как эталонного, так и исследуемого сигнала. Это внутреннее представление служит оценкой информации, доступной человеческому мозгу для сравнения сигналов. Данный метод называется сравнением внутренних представлений и используется в ASD.

3 Психоакустические и когнитивные основы

В этом разделе обсуждаются свойства слуховой системы человека, наиболее важные при оценке воспринимаемого качества аудиосигналов. Основное внимание уделяется способам моделирования этих свойств.

РИСУНОК 5

Психоакустические концепции, используемые в различных схемах перцептивного измерения



Сравнение внутренних представлений

Концепция порога маскирования

BS.1387-05

3.1 Передаточная характеристика наружного и среднего уха

Как правило, звуковые сигналы должны пройти через наружное и среднее ухо, пока не достигнут внутреннего уха, где выполняются процессы улавливания и анализа звука. Наружное и среднее ухо выполняют функции полосового фильтра входного сигнала. К входному сигналу добавляется шум, присутствующий в слуховом нерве, вместе с шумом, вызванным потоком крови. Амплитуда этого шума выше на низких частотах. Передаточная функция наружного и среднего уха вместе с внутренним шумом ограничивают способность улавливать слабые звуковые сигналы и оказывают наибольшее влияние на абсолютный порог слышимости.

3.2 Шкалы воспринимаемых частот

Рецепторами звукового давления в человеческом ухе служат волосковые сенсорные клетки. Они расположены во внутреннем ухе, точнее в улитке. В улитке выполняется преобразование частоты в пространственное положение. Положение максимального возбуждения зависит от частоты входного сигнала. Каждая волосковая сенсорная клетка в определенном месте улитки отвечает за определенный перекрывающийся диапазон на шкале частот. Создаваемое впечатление высоты звука коррелирует с постоянным расстоянием между волосковыми сенсорными клетками.

В ходе разнообразных психоакустических экспериментов были обнаружены различные функции преобразования частоты в высоту тона.

В таблице, приведенной в [Zwicker and Feldtkeller, 1967], шкала частот в герцах разбита на 24 непересекающиеся полосы, так называемые критические полосы. Верхние граничные частоты этих полос приведены в таблице 5. В этой таблице также дано определение шкалы Барка: 1 барк соответствует 100 Гц, 24 барка соответствуют 15 500 Гц.

ТАБЛИЦА 5

Шкала критических полос согласно определению Цвикера

Критическая полоса	1	2	3	4	5	6	7	8	9	10	11	12
Верхняя граничная частота (Гц)	100	200	300	400	510	630	770	920	1080	1270	1480	1720

Критическая полоса	13	14	15	16	17	18	19	20	21	22	23	24
Верхняя граничная частота (Гц)	2000	2320	2700	3150	3700	4400	5300	6400	7700	9500	12 000	15 500

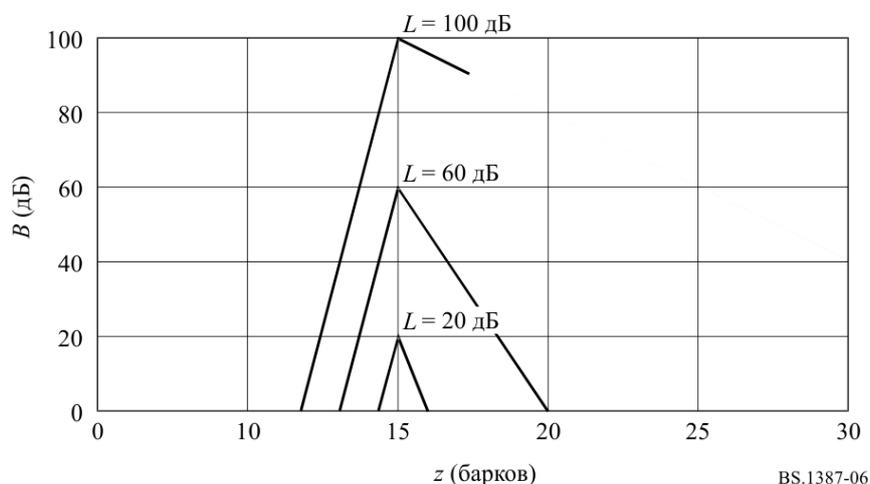
В прошлом появилось несколько приближений к шкале Барка. Подробное обсуждение различных шкал содержится в [Cohen and Fielder, 1992]. В контексте объективного измерения воспринимаемого качества звука наилучшие результаты были достигнуты с использованием шкалы Барка.

3.3 Возбуждение

Каждая волосковая сенсорная клетка реагирует на диапазон частот, который можно описать характеристикой фильтра. Наклон характеристики фильтра лучше всего можно выразить на шкале восприятия, как описано выше. Форма характеристики фильтров на такой шкале практически не зависит от центральной частоты. Наклон нижней части кривой возбуждения не зависит от уровня L входного сигнала (около 27 дБ/барк). Наклон в верхней части более крутой для низких уровней входного сигнала, чем для высоких (от -5 до -30 дБ/барк). Эта крутая характеристика вызвана механизмом обратной связи между двумя разными типами волосковых сенсорных клеток, и для ее стабилизации требуется некоторое время. Поэтому наилучшее частотное разрешение слуха для стационарных сигналов достигается через несколько миллисекунд после начала сигнала. Шаблоны возбуждения сигналов, состоящих из нескольких составляющих, складываются нелинейным образом.

РИСУНОК 6

Зависимости уровней возбуждения по Терхардту [1979]



BS.1387-06

После воздействия сигнала волосковым сенсорным клеткам и нервной системе требуется некоторое время для восстановления, пока вновь не будет достигнута полная чувствительность. Длительность процесса восстановления зависит от уровня и длительности сигнала и может достигать нескольких сотен миллисекунд. Высокоуровневые сигналы обрабатываются на пути между волосковыми сенсорными клетками и мозгом быстрее, чем низкоуровневые. Следовательно, начало громкого сигнала может маскировать предшествующий более тихий сигнал.

Другой подход к моделированию возбуждения основан на шкале ERB [Moore, 1986]. При этом подходе используются так называемые фильтры ROEX [Moore, 1986]. В контексте объективного измерения воспринимаемого качества звука лучшие результаты были достигнуты с помощью моделей, основанных на работах [Zwicker and Feldtkeller, 1967] и [Terhardt, 1979].

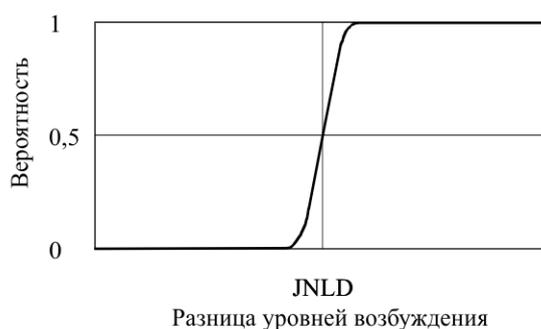
3.4 Обнаружение

Возбуждение, вызванное различными звуковыми сигналами, передается в мозг. Выделяют три вида памяти, которые различаются по степени детализации и продолжительности хранения информации: долговременная, кратковременная и сверхкратковременная. В контексте испытаний с прослушиванием наиболее важную роль играет сверхкратковременная память. Большинство деталей сигнала сохраняется, если продолжительность аудиофрагмента составляет менее пяти-восьми секунд в зависимости от слушателя и аудиофрагмента. Это учитывается в процедуре оценки, определенной в Рекомендации МСЭ-R BS.1116, где испытуемым разрешается выбирать очень короткие части аудиофрагмента для более внимательного прослушивания. На пороге обнаружения вероятность обнаружения составляет 50%. Вокруг этого порога вероятность обнаружения разницы плавно возрастает от 0% до 100%.

Порогом обнаружения разницы уровней служит едва заметная разница уровней (Just-Noticeable Level Difference – JNLD). JNLD зависит от уровня входных сигналов. Для обнаружения слабых сигналов необходима большая разница (уровень 20 дБ SPL, JNLD 0,75 дБ). Для громких сигналов чувствительность к небольшой разнице уровней намного выше (уровень 80 дБ SPL, JNLD 0,2 дБ). Эти цифры определены по результатам экспериментов по амплитудной модуляции.

РИСУНОК 7

Принцип вероятности обнаружения



BS.1387-07

3.5 Маскирование

Сигнал, ясно слышимый, когда он подается отдельно, может быть совершенно неслышимым в присутствии другого сигнала – маскирующего. Этот эффект называется маскированием. Следует различать две ситуации.

– Одновременное маскирование

В этой ситуации маскирующий и маскируемый сигналы присутствуют одновременно и являются квазистационарными. Если маскирующий сигнал имеет дискретную ширину полосы, то порог слышимости повышается даже для частот ниже или выше частоты маскирующего сигнала. Степень маскирования зависит от структуры маскирующего и маскируемого сигналов. В ситуации, когда шумоподобный сигнал маскирует тональный сигнал, степень маскирования практически не зависит от частоты. Если уровень звукового давления маскируемого сигнала примерно на 5 дБ ниже уровня маскирующего сигнала, он становится неслышимым. В ситуации, когда тональный сигнал маскирует шумоподобный сигнал, степень маскирования зависит от частоты маскирующего сигнала. Ее можно оценить по формуле $\left(15,5 + \frac{z}{\text{барк}}\right)$ дБ, где z – критическая полоса частот маскирующего сигнала.

Кроме того, при высоких уровнях сигнала нелинейные эффекты уменьшают порог маскирования вблизи полосы маскирующего сигнала. Аналогичные эффекты возникают при маскировании одного тона другим. Пороги маскирования нескольких сигналов суммируются нелинейным образом. Обычно результирующий порог маскирования превышает порог маскирования, создаваемый каждым отдельным сигналом.

– *Временное маскирование*

В этой ситуации маскирующий и маскируемый сигналы подаются в разное время. Вскоре после спада маскирующего сигнала порог маскирования оказывается ближе к порогу одновременного маскирования этого сигнала, чем к абсолютному порогу. В зависимости от продолжительности маскирующего сигнала время спада порога может составлять от 5 мс (маскирующий сигнал – гауссов импульс длительностью около 0,05 мс) до 150 мс и более (маскирующий сигнал – розовый шум длительностью 1 с). Слабые сигналы маскируются непосредственно перед более громкими. Продолжительность этого эффекта обратного маскирования составляет около 5 мс. Если маскируемый сигнал находится чуть выше порога, он воспринимается не перед маскирующим сигналом, а как его изменение. Эффект обратного маскирования демонстрирует значительные отклонения от слушателя к слушателю.

3.6 Громкость и частичное маскирование

Воспринимаемая громкость аудиосигналов зависит от их частоты, длительности и уровня звукового давления. Благодаря автомаскированию громкость сложного сигнала оказывается меньше суммарной громкости всех его составляющих. В контексте измерения качества звука громкость нежелательных искажений, добавляемых к эталонному сигналу, то есть громкость шума, снижается за счет частичного маскирования, обусловленного эталонным сигналом.

3.7 Резкость звука

Резкость звука – одна из основных характеристик восприятия – связана с тембром. Звук воспринимается как резкий, если он содержит преимущественно высокочастотные составляющие. Например, резким считается шум синусоидального тона, или высокочастотный шум с ограниченной полосой, либо шум в полосе верхних частот с частотой среза в диапазоне частот выше примерно 3 кГц. Однако точная частотная характеристика аудиосигнала не оказывает существенного влияния на резкость звука. Фундаментальное исследование резкости звука было проведено фон Бисмарком [1974].

Дополнительные исследования резкости звука описаны в [Aures, 1984]. Результатом этих исследований стала несколько измененная весовая функция по сравнению с весовой функцией, определенной Бисмарком. Она в меньшей степени способствует определению звука как резкого на очень низких и очень высоких критических частотах и в большей – при критических частотах между 14 и 20 барками. Кроме того, эти исследования показали, что резкость аудиосигналов с большими колебаниями уровня звукового давления и сильным высокочастотным содержанием определяется не только общей громкостью, но и весовой функцией, зависящей от общей громкости.

3.8 Когнитивная обработка

Понятно, что воспринимаемое качество звука в значительной степени зависит от когнитивных эффектов. Это можно продемонстрировать с помощью простого эксперимента.

Эталонный сигнал с отчетливо слышимым фоновым шумом обрабатывается неким аудиооборудованием, не способным передавать этот фоновый шум. Поскольку шум представляет собой нежелательное искажение, при испытании с прослушиванием эталонный сигнал получит более низкую оценку, чем обработанный. С другой стороны, тот же обработанный сигнал получил бы худшие оценки, если бы наиболее важной частью эталонного сигнала был тихий фоновый шум.

Перечисление всех возможных когнитивных эффектов выходит за рамки данного Прилагаемого документа, но ниже приводятся некоторые примеры.

Пример 1. Отделение линейных искажений от нелинейных

Линейные искажения менее неприятны, чем нелинейные. Отделение линейных искажений от нелинейных можно довольно легко реализовать с помощью адаптивной обратной фильтрации выходного сигнала. Отделение линейных искажений от нелинейных используется в методике, приведенной в настоящей Рекомендации.

Пример 2. Анализ акустической картины

Анализ акустической картины [Bregman, 1990] – это когнитивный процесс, позволяющий слушателям разделять различные акустические события и группировать их в разные образы. Для количественной оценки эффекта анализа акустической картины оказался полезным прагматический подход, предложенный в [Beerends and Stemerdink, 1994]. Если частотно-временная составляющая не кодируется кодеком, то оставшийся сигнал по-прежнему образует единую связную акустическую картину, тогда как введение новой, несвязанной частотно-временной составляющей приводит к ощущению двух отдельных сигналов. Ввиду разделения на два отдельных восприятия искажение будет более неприятным, чем можно было бы ожидать, исходя из громкости вновь введенного искажающего компонента. Это приводит к асимметрии между воспринимаемым искажением, вызванным отсутствием кодирования частотно-временного компонента, и искажением, вызванным введением нового частотно-временного компонента.

Пример 3. Информационное маскирование

Информационное маскирование можно смоделировать, определив энтропийный показатель спектрально-временной сложности. Этот эффект, скорее всего, зависит от количества учебных занятий, которые проходят испытуемые перед проведением субъективной оценки. Первая попытка смоделировать этот эффект сделана в [Beerends *et al*, 1996], где рассчитывается оценка локальной сложности во временном окне около 100 мс. Если эта локальная сложность высока, то искажения в этом временном окне услышать труднее, чем при низкой локальной сложности. Обучение может снизить порог маскирования на несколько десятков децибел [Leek and Watson, 1984].

Пример 4. Спектрально-временное взвешивание

Некоторые спектрально-временные области аудиосигнала несут больше информации и поэтому могут быть более важными, чем другие. Обнаружено, что спектрально-временное взвешивание важно для оценки качества речевых кодеков. В речи одни спектрально-временные компоненты, например форманты, явно несут в себе больше информации, чем другие [Beerends and Stemerdink, 1994]. Однако в музыке информацию могут нести все спектрально-временные компоненты сигнала, даже паузы.

4 Встроенные модели

4.1 DIX

Метод измерения восприятия DIX (индекс искажений) [Thiede and Kabot, 1996] основан на наборе слуховых фильтров, который обеспечивает высокое временное разрешение и, таким образом, позволяет более точно (по сравнению с подходами на основе БПФ) моделировать временные эффекты, такие как пре- и постмаскирование. Временная тонкая структура огибающих на каждом слуховом фильтре сохраняется и используется для получения дополнительной информации о сигналах и внесенных искажениях.

Центральные частоты отдельных фильтров равномерно распределены по шкале воспринимаемой высоты тона. Верхняя часть формы частотной характеристики фильтра слегка закруглена, чтобы выбранное количество фильтров охватывало весь частотный диапазон без пульсаций общей частотной характеристики. Чтобы смоделировать пороги маскирования, наклон частотных характеристик фильтра экспоненциально уменьшается вдоль шкалы Барка. Крутизна наклона частотных характеристик фильтра зависит от уровня входных сигналов. В первой версии DIX диапазон слышимых частот покрывался 80 фильтрами, а затем их число было сокращено до 40 фильтров, то есть разрешение по частоте соответствует примерно 0,6 барка. Этот алгоритм фильтрации с использованием набора фильтров работает довольно быстро по сравнению с другими наборами

фильтров, состоящими из отдельных фильтров, но все же требует гораздо больше времени, чем блочные преобразования, такие как БПФ и вейвлет-преобразования пакетов.

DIX динамически адаптирует уровни и спектры между исследуемым и эталонным сигналами, отделяя линейные искажения от нелинейных. Он оценивает структуру временных огибающих на выходе фильтра, чтобы смоделировать повышенную степень маскирования, вызванную модулированными и шумоподобными маскирующими устройствами, по сравнению с чистыми тонами.

Многочисленные выходные параметры, включая парциальную громкость нелинейных искажений, показатели количества линейных искажений и меры временных и бинауральных эффектов рассчитываются путем сравнения внутренних представлений исследуемого и эталонного сигналов. Однако хорошей оценки базового качества звука можно добиться, используя всего два выходных параметра – преобразуя парциальную громкость нелинейных искажений вместе с одним из показателей количества линейных искажений в оценку ожидаемого базового качества звука исследуемого сигнала.

4.2 NMR

Схема измерения NMR (отношение уровня шума к уровню маскируемого сигнала) [Brandenburg, 1987] оценивает разницу уровней между порогом маскирования и шумовым сигналом. Для анализа частотного состава сигнала используется ДПФ с окном Ханна около 20 мс. Коэффициенты преобразования объединяются в полосы в соответствии со шкалой Барка. Для каждой полосы оценивается порог маскирования. Наклон кривой порога маскирования определяется с использованием подхода наихудшего случая, принимая во внимание тот факт, что для слабых сигналов наклон круче, а при более высоких уровнях достигается абсолютный порог. Абсолютный порог адаптирован не к психоакустическим требованиям, а к разрешению входного сигнала (обычно 16 битов). Благодаря этим фактам NMR устойчиво к изменениям уровня воспроизведения. Разрешение шкалы наклона составляет около 1 барка. Поскольку требуемая вычислительная мощность невелика, на ранней стадии разработки удалось реализовать NMR как систему реального времени.

Модель находится в эксплуатации с 1987 года и доказала свою принципиальную надежность.

Наиболее важными выходными значениями NMR являются частота флагов маскирования, указывающая процент кадров со слышимыми искажениями, а также общее и среднее значение NMR, которые соответствуют разным способам усреднения интервала между энергией ошибки и порогом маскирования.

4.3 OASE

В схеме измерения объективной оценки аудиосигнала (Objective Audio Signal Evaluation – OASE) [Sporeg, 1997] для анализа входных сигналов используется набор из 241 фильтра. Центральные частоты равномерно распределены по шкале Барка с шагом 0,1 барка. Фильтры перекрывают друг друга. Каждый из фильтров адаптирован к частотной характеристике точки на базилярной мембране. Зависимость наклона характеристик от уровня учитывается с помощью подхода наихудшего случая, как это делается в NMR. Для фильтров с низкими центральными частотами требуется расчет с полной частотой дискретизации, а для фильтров с более высокими центральными частотами может производиться расчет с уменьшенной частотой дискретизации. После применения фильтров производится расчет модели временных эффектов слуховой системы человека, как в ASD. После этого возможно снижение частоты дискретизации во всех полосах фильтрации. Это приводит к временному разрешению набора фильтров 0,66 мс при частоте дискретизации 48 кГц. Выходные данные согласующих фильтров эталонного и исследуемого сигналов сравниваются с функцией вероятности обнаружения. Эта функция использует громкость входных сигналов в качестве входных данных для расчета JNLD. Общая вероятность обнаружения складывается из вероятностей обнаружения для каждой полосы частот. Эта операция выполняется для обоих входных каналов, а также для так называемого центрального канала. Вероятность обнаружения для центрального канала каждой полосы частот представляет собой наихудший случай вероятности обнаружения для левого и правого каналов. Для каждого кадра длительностью 0,66 мс также рассчитывается сумма шагов выше порога.

Используется несколько способов усреднения вероятности обнаружения и шагов выше порога по времени:

- усредненное по времени значение вероятности обнаружения;
- частота кадров с вероятностью обнаружения выше 0,5;
- максимальная вероятность обнаружения с фильтром нижних частот;
- максимальная вероятность обнаружения с фильтром нижних частот с забыванием;
- среднее количество шагов выше порога для кадров с вероятностью обнаружения выше 0,5;
- среднее количество шагов выше порога;
- максимальное количество шагов выше порога;
- среднее количество шагов выше порога для 10% наихудших кадров.

4.4 Показатель воспринимаемого качества звука (PQM)

Основной принцип PQM [Beerends and Stemerding, 1992] заключается в определении разницы между внутренними представлениями (в голове субъекта) эталонного и искаженного сигналов и ее преобразовании в субъективно воспринимаемое качество звука по когнитивному принципу. Преобразование из физической, внешней сферы в психофизическую, внутреннюю сферу осуществляется посредством четырех операций:

- частотно-временное преобразование посредством ДПФ с окном Ханна длительностью около 40 мс;
- частотное искажение с использованием шкалы Барка;
- частотно-временное расширение спектра (нелинейная свертка);
- искажение по интенсивности (сжатие).

Сочетание размытия и сжатия позволяет моделировать маскирующее поведение слуховой системы человека на пороге маскирования и выше. Оптимизация сжатия выполняется с использованием субъективных результатов первой оценки аудиокодека MPEG [ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG90/N0030, October 1990], [ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG91/N0010, June 1991]. Разница во внутреннем представлении выражается через шумовые помехи. В последних версиях PQM, представленных в МСЭ-R, в процесс преобразования шумовых помех в субъективное качество были включены два когнитивных эффекта: воспринимаемый поток [Beerends и Stemerding, 1994] и информационное маскирование [Beerends *et al*, 1996].

С использованием когнитивной модели, представленной в [Beerends and Stemerding, 1994], была разработана упрощенная версия PQM – показатель воспринимаемого качества речи (Perceptual Speech Quality Measure – PSQM) [Beerends and Stemerding, 1994], – дополненная взвешиванием интервалов молчания. В ходе разработки PSQM выяснилось, что при оценке качества речи в контексте телефонии шум, возникающий во время интервалов молчания, имеет меньшее значение, чем шум, возникающий во время интервалов активной речи. В ходе испытаний МСЭ-Т предложение PSQM показало самую высокую корреляцию между объективным и субъективным качеством (12-я Исследовательская комиссия по стандартизации электросвязи, COM 12-74 – Обзор проверочных испытаний объективных показателей качества речи). Оценка воспринимаемого качества речи (PESQ) с улучшенным алгоритмом оценки определена в Рекомендации МСЭ-Т P.862.

4.5 Алгоритм PERCEVAL

Алгоритм PERCEVAL (PERceptual EVALuation) [Paillard *et al*, 1992] моделирует характеристики передачи среднего и внутреннего уха в целях формирования внутреннего представления сигнала. Входной сигнал разлагается на частотно-временное представление с помощью ДПФ. Обычно к входным данным применяется окно Ханна длительностью около 40 мс с 50%-ным перекрытием между последовательными окнами. Энергетический спектр умножается на частотно-зависимую функцию, моделирующую воздействие слухового прохода и среднего уха. Значения ослабленной спектральной энергии преобразуются из шкалы частоты в шкалу высоты тона, которая более линейна

как по отношению к физическим свойствам внутреннего уха, так и по отношению к наблюдаемым психофизическим эффектам. Затем преобразованные энергетические компоненты свертываются с помощью функции расширения спектра для имитации рассеивания энергии вдоль базилярной мембраны. Наконец, к каждому компоненту высоты звука добавляется собственная частотно-зависимая энергия для учета абсолютного порога слышимости. Преобразование энергии в децибелы приводит к представлению сигнала на базилярной мембране.

При моделировании экспериментов по слуховому маскированию для каждого входного сигнала формируется представление на базилярной мембране, а информация о разнице между этими представлениями используется для решения задачи. Одно представление состоит только из маскирующего сигнала, а другое представляет собой комбинацию маскирующего и исследуемого сигналов. Их разность соответствует незамаскированной составляющей сигнала. Алгоритм PERCEVAL вычисляет вероятность обнаружения этой разности. Вероятность необнаружения разности для каждого детектора вдоль моделируемой базилярной мембраны оценивается с помощью сигмоидальной функции вероятности. В предположении, что детекторы статистически независимы, глобальная вероятность обнаружения полного набора детекторов рассчитывается как дополнение произведения отдельных вероятностей необнаружения. С помощью этого подхода было успешно смоделировано несколько экспериментов по маскированию; данная модель также использовалась для оценки возможности моделирования отдельных слушателей [Treurniet, 1996].

В качестве инструмента для оценки качества звука PERCEVAL вычисляет разность между представлениями эталонного и исследуемого сигналов. Применяя разумные предположения о перцептивных и когнитивных процессах более высокого уровня, вычисляется ряд перцептивно значимых переменных и сопоставляется с показателем объективного качества исследуемого сигнала. Преобразование было оптимизировано путем минимизации разницы между распределением объективного качества и соответствующим распределением средних субъективных оценок качества для имеющегося набора данных.

4.6 Модель POM

Воспринимаемый объективный показатель (Perceptual Objective Measurement – POM) [Colomes *et al*, 1995] дает количественную оценку определенного уровня ухудшения сигнала, которое может наблюдаться между эталонным сигналом и его ухудшенной версией. Это достигается путем сравнения внутреннего базилярного представления обоих сигналов, чем бы ни было вызвано ухудшение. Базилярное представление моделирует различные процессы, которым подвергается аудиосигнал при прохождении через человеческое ухо. Следовательно, первым этапом POM является вычисление внутреннего представления аудиосигнала. Для моделирования частоты импульсов в нейронах вдоль базилярной мембраны выбран шаблон возбуждения (представленного в децибелах), распространяющегося по базилярной мембране.

Процесс расчета шаблона возбуждения называется искусственным ухом. Затем, когда имеется два внутренних представления сигналов, которые нужно сравнить друг с другом, POM должен определить, слышна ли разница между этими внутренними представлениями, и если да, то какая. Это называется процессом обнаружения.

POM использует ДПФ с окном Ханна длительностью примерно 40 мс (с 50%-ным перекрытием между двумя окнами Ханна). Количество базилярных каналов анализа – 620. Остальные части слуховой модели практически идентичны тем, которые используются в обеих моделях RAQM и PERCEVAL.

Функция расширения спектра достаточно точно описывается более точной аппроксимацией, учитывающей как зависимость от уровня согласно [Terhardt, 1979], так и округлую форму согласно [Schroeder *et al*, 1979].

Эта модель дает вероятность обнаружения искажения между двумя сравниваемыми сигналами, а также так называемое базилярное расстояние, которое представляет собой воспринимаемый разрыв между двумя сравниваемыми возбуждениями.

4.7 Инструментальный метод

Инструментальный метод использует трехэтапный подход для измерения воспринимаемой разницы между качеством звука исследуемого и эталонного аудиосигналов, что дает представление об общем субъективном уровне качества звука исследуемого сигнала. Метод основан на хорошо известных моделях восприятия, которые используются для описания воспринимаемого представления о различиях между двумя аудиосигналами. Кроме того, он включает в себя процедуру взвешивания воспринимаемого качества звука исследуемого стереосигнала с учетом результатов как для левого, так и для правого каналов. Жесткая корреляция по каждой выборке эталонного и исследуемого аудиосигналов не требуется.

Основная функция инструментального метода на первом этапе базируется на вычислении удельной громкости, рассчитанной согласно [Zwicker and Feldtkeller, 1967] с использованием БПФ из 2048 точек с окном Ханна, соответствующим по продолжительности примерно 40 мс. Все окно сдвигается с шагом 10 мс. Кроме того, применяются эффекты временного маскирования по Цвикеру, такие как пост- и предмаскирование. На основании этих базовых параметров ощущения в качестве предварительной обработки для следующих этапов вычисляются другие параметры восприятия, такие как интегральная громкость, частично маскированная громкость, резкость согласно [von Bismarck, 1974] и [Aures, 1984], а также количество опережающих эхо.

На втором этапе инструментального метода выполняются процедуры взвешивания, зависящие главным образом от величины воспринимаемой разницы в громкости и изменений громкости во времени.

На третьем этапе инструментальный метод генерирует набор промежуточных выходных значений, основанных на статистическом анализе результатов первого и второго этапов. Результаты этого статистического анализа содержат средние, максимальные и среднеквадратичные значения, а также стандартное отклонение средних значений. Взвешенная сумма этих промежуточных выходных данных инструментального метода используется для окончательного определения воспринимаемой разницы между исследуемым и эталонным сигналами. При необходимости это единственное выходное значение можно сопоставить с субъективной оценкой разницы, обычно получаемой в ходе субъективных испытаний с прослушиванием путем последовательной подгонки выходных данных для каждого приращения времени с использованием линейной функции или полиномиальной функции более высокого порядка.

Приложение 2

Описание модели²

1 Основные принципы

РИСУНОК 8
Общая блок-схема системы измерений



BS.1387-08

Предлагаемая *Методика объективного измерения воспринимаемого качества звука* включает в себя модель периферического уха, несколько промежуточных этапов (называемых здесь *предварительной обработкой шаблонов возбуждения*), психоакустический (в основном) расчет *выходных переменных модели* (MOV) и преобразование набора выходных переменных модели в единое значение, представляющее *базовое качество звука* исследуемого сигнала. Она включает в себя две модели периферического уха – на основе БПФ и на основе набора фильтров. Общая структура обеих моделей периферического уха одинакова, за исключением расчета сигнала ошибки (который используется только в модели уха на основе БПФ).

² Приверженцы технологии, описанной в этой Рекомендации, подали заявки на патенты, соответствующие Приложению 1 Резолюции МСЭ-R 1. Технология, описанная в настоящей Рекомендации, защищена международными патентами и, как и все Рекомендации МСЭ, авторским правом. Для использования этой технологии необходимо предварительное согласие владельцев в форме лицензии. За дополнительной информацией по лицензированию этой технологии просьба обращаться к патентной базе данных МСЭ-R или в Секретариат БР.

Входные данные для расчета MOV:

- шаблоны возбуждения как для исследуемого, так и для эталонного сигналов;
- спектрально адаптированные шаблоны возбуждения как для исследуемого, так и для эталонного сигналов;
- шаблоны удельной громкости как для исследуемого, так и для эталонного сигналов;
- шаблоны модуляции как для исследуемого, так и для эталонного сигналов;
- сигнал ошибки, рассчитываемый как спектральная разность между исследуемым и эталонным сигналами (только для модели уха на основе БПФ).

Если не указано иное, то в случае стереосигналов все вычисления выполняются независимо и одинаково для левого и правого каналов.

В описании определены две установки, называемые *базовой версией* и *расширенной версией*.

Во всех приведенных уравнениях обозначение Ref. используется для всех шаблонов, рассчитанных на основе эталонного сигнала, а Test – для всех шаблонов, рассчитанных на основе исследуемого сигнала. Индекс k указывает на дискретную переменную частоты (то есть полосу частот), а n – на дискретную переменную времени (то есть либо число кадров, либо число выборок). Если значения k или n не определено явно, то вычисления выполняются для всех возможных значений k и n . Все остальные сокращения поясняются там, где они встречаются.

В названиях выходных переменных модели индекс A используется для всех переменных, рассчитанных с использованием модели уха, основанной на наборе фильтров, а индекс B – для всех переменных, рассчитанных с использованием модели уха, основанной на БПФ.

1.1 Базовая версия

Базовая версия включает только те MOV, которые рассчитаны с использованием модели уха на основе БПФ. Модель, основанная на наборе фильтров, не используется. В *базовой версии* используется в общей сложности 11 MOV для прогнозирования воспринимаемого *базового качества звука*.

1.2 Расширенная версия

Расширенная версия включает MOV, рассчитанные с использованием модели уха на основе набора фильтров, а также MOV, рассчитанные с использованием модели уха на основе БПФ. Спектрально адаптированные шаблоны возбуждения и шаблоны модуляции вычисляются только с использованием модели, основанной на наборе фильтров. В *расширенной версии* используется в общей сложности 5 MOV для прогнозирования воспринимаемого *базового качества звука*.

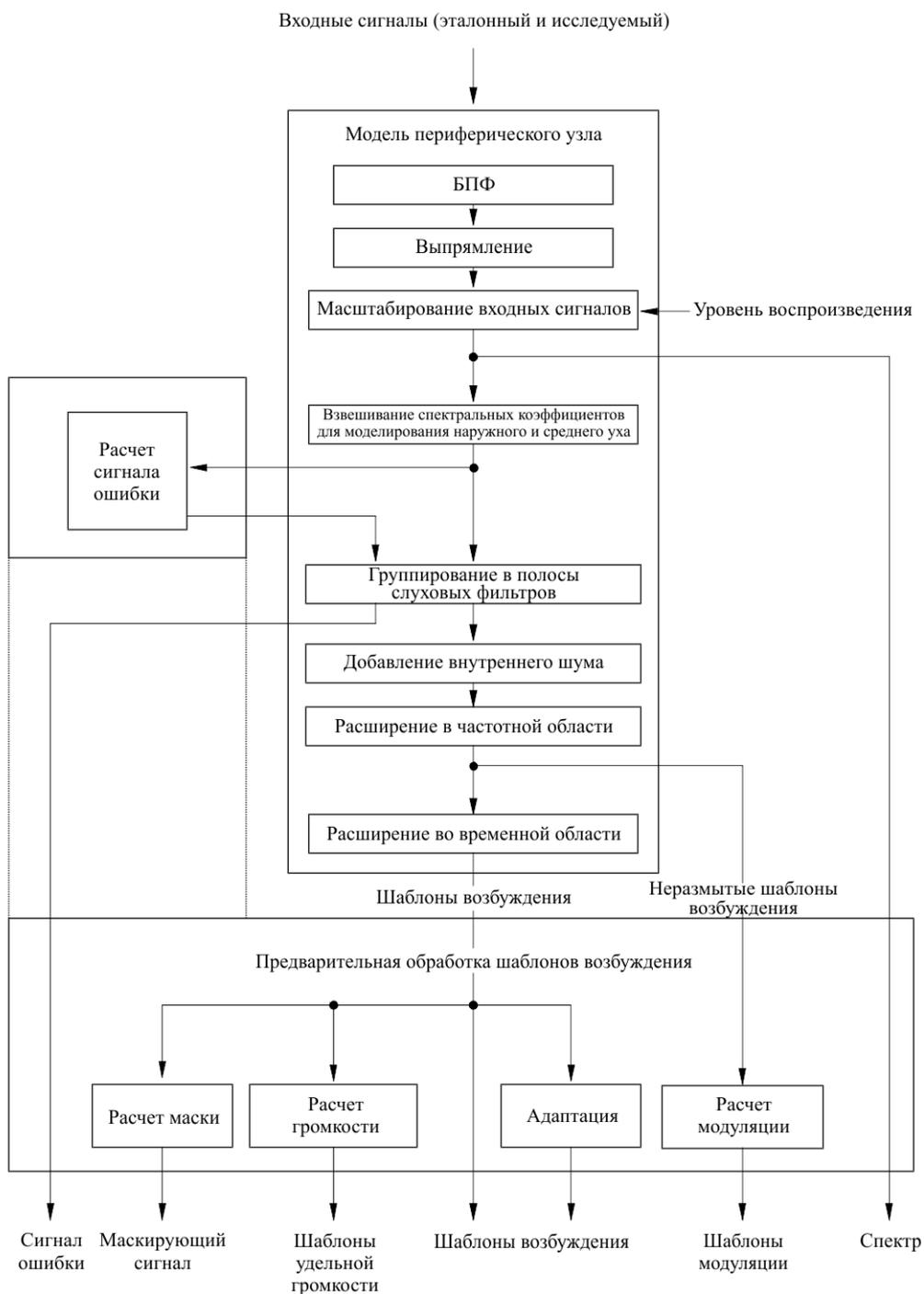
2 Модель периферического уха

2.1 Модель уха на основе БПФ

2.1.1 Обзор

РИСУНОК 9

Модель периферического уха и предварительная обработка шаблонов возбуждения для модели, основанной на БПФ



Входные данные модели уха на основе БПФ – эталонный и исследуемый сигналы с частотой дискретизации 48 кГц, синхронизированные по времени, – разделяются на кадры длительностью около 0,042 с с перекрытием 50%. Каждый кадр преобразуется в частотную область с использованием окна Ханна и краткосрочного БПФ и масштабируется до уровня воспроизведения. К спектральным коэффициентам применяется весовая функция, моделирующая частотную характеристику наружного и среднего уха. Преобразование в представление основного тона осуществляется путем группирования взвешенных спектральных коэффициентов в критические полосы. Для имитации внутреннего шума в слуховой системе добавляется частотно-зависимое смещение. Для моделирования спектральных слуховых фильтров в частотной области используется функция расширения спектра, зависящая от уровня. Она следует за расширением во временной области, которое учитывает эффекты прямого маскирования.

Полученные *шаблоны возбуждения* используются для расчета *шаблонов удельной громкости* и *шаблонов маскирования*. Шаблоны, полученные до окончательного расширения во временной области (*неразмытые шаблоны возбуждения*), используются для расчета *шаблонов модуляции*.

Чтобы смоделировать сигнал ошибки, шаблоны эталонного и исследуемого сигналов на выходе фильтра наружного и среднего уха объединяются и отображаются на шкале высоты тона путем группирования по критическим полосам.

Эти выходные данные используются вместе с шаблонами возбуждения для расчета значений выходных переменных модели.

2.1.2 Обработка по времени

Входные данные модели уха на основе БПФ – эталонный и исследуемый сигналы – разделяются на кадры по 2048 выборок с перекрытием в 1024 выборках:

$$t_n[k_t, n] = t[1024 \cdot n + k_t], \quad n = 0, 1, 2, \dots, k_t = 0 \dots 2047, \quad (1)$$

где n – номер временного кадра, а k_t – номер интервалов времени внутри кадра.

2.1.3 БПФ

Преобразование из временной области в частотную осуществляется с использованием окна Ханна:

$$h_w[k] = \frac{1}{2} \sqrt{\frac{8}{3}} \left[1 - \cos\left(2\pi \frac{k}{N-1}\right) \right] \Big|, \quad N = 2048; \quad (2)$$

$$t_w[k_t, n] = h_w[k_t] \cdot t_n[k_t, n]. \quad (3)$$

Затем следует кратковременное преобразование Фурье:

$$F_f[k_f, n] = \frac{1}{2048} \sum_{k_t=0}^{2047} t_w[k_t, n] e^{-j \frac{2\pi}{2048} k_f k_t}. \quad (4)$$

Коэффициент масштабирования БПФ рассчитывается на основе предполагаемого уровня звукового давления L_p полномасштабной синусоидальной волны по формуле

$$fac = \frac{10^{\frac{L_p}{20}}}{Norm}; \quad (5)$$

$$F[k_f, n] = fac \cdot F_f[k_f, n]. \quad (6)$$

Коэффициент нормализации $Norm$ рассчитывается путем принятия в качестве входного сигнала синусоидального сигнала с частотой 1019,5 Гц и полной шкалой 0 дБ и расчета максимального абсолютного значения спектральных коэффициентов по 10 кадрам.

Если уровень звукового давления неизвестен, рекомендуется установить L_p равным 92 дБ_{SPL}.

2.1.4 Наружное и среднее ухо

Частотная характеристика наружного и среднего уха моделируется частотно-зависимой весовой функцией:

$$W[k]/\text{дБ} = -0,6 \cdot 3,64 \cdot \left(\frac{f[k]}{\text{кГц}}\right)^{-0,8} + 6,5 \cdot e^{-0,6 \cdot \left(\frac{f[k]}{\text{кГц}} - 3,3\right)^2} - 10^{-3} \cdot \left(\frac{f[k]}{\text{кГц}}\right)^{3,6}, \quad (7)$$

где:

$$f[k]/\text{Гц} = k \cdot 23,4375 \quad (8)$$

есть частотное представление в строке k , которое применяется к выходным данным БПФ (уравнение (9)):

$$F_e[k_f, n] = \left| F[k_f, n] \right| \cdot 10^{\frac{W[k_f]}{20}}. \quad (9)$$

Величина $F_e[k_f]$ называется *выходными данными БПФ, взвешенными по наружному уху*.

2.1.5 Группирование в критические полосы

Шкала слухового тона рассчитывается на основе приближения, предложенного в [Schroeder *et al*, 1979]:

$$z/\text{барк} = 7 \cdot \operatorname{arsinh} \left(\frac{f/\text{Гц}}{650} \right). \quad (10)$$

Единицей высоты тона служит *барк* (хотя эта шкала не совсем соответствует шкале Барка, определенной в [Zwicker and Feldtkeller, 1967]).

Частотные границы фильтров находятся в диапазоне от 80 Гц до 18 000 Гц. Ширина и расстояние между полосами фильтра соответствуют разрешению $res = 0,25$ барка для базовой версии и $res = 0,5$ барка для расширенной версии.

Это дает количество частотных полос $Z = 109$ для базовой версии и $Z = 55$ для расширенной версии.

ТАБЛИЦА 6

Полосы частот модели уха на основе БПФ, используемой в базовой версии

Группа	Нижняя частота/Гц	Центральная частота/Гц	Верхняя частота/Гц	Ширина полосы/Гц
k	$f_l[k]$	$f_c[k]$	$f_u[k]$	$f_w[k]$
0	80	91,708	103,445	23,445
1	103,445	115,216	127,023	23,577
2	127,023	138,87	150,762	23,739
3	150,762	162,702	174,694	23,932
4	174,694	186,742	198,849	24,155
5	198,849	211,019	223,257	24,408
6	223,257	235,566	247,95	24,693
7	247,95	260,413	272,959	25,009
8	272,959	285,593	298,317	25,358
9	298,317	311,136	324,055	25,738
10	324,055	337,077	350,207	26,151
11	350,207	363,448	376,805	26,598
12	376,805	390,282	403,884	27,079
13	403,884	417,614	431,478	27,594
14	431,478	445,479	459,622	28,145
15	459,622	473,912	488,353	28,731
16	488,353	502,95	517,707	29,354
17	517,707	532,629	547,721	30,014
18	547,721	562,988	578,434	30,713
19	578,434	594,065	609,885	31,451
20	609,885	625,899	642,114	32,229
21	642,114	658,533	675,161	33,048
22	675,161	692,006	709,071	33,909
23	709,071	726,362	743,884	34,814
24	743,884	761,644	779,647	35,763
25	779,647	797,898	816,404	36,757
26	816,404	835,17	854,203	37,799
27	854,203	873,508	893,091	38,888
28	893,091	912,959	933,119	40,028
29	933,119	953,576	974,336	41,218
30	974,336	995,408	1016,797	42,461
31	1016,797	1038,511	1060,555	43,758
32	1060,555	1082,938	1105,666	45,111
33	1105,666	1128,746	1152,187	46,521
34	1152,187	1175,995	1200,178	47,991

ТАБЛИЦА 6 (продолжение)

Группа	Нижняя частота/Гц	Центральная частота/Гц	Верхняя частота/Гц	Ширина полосы/Гц
k	$f_l[k]$	$f_c[k]$	$f_u[k]$	$f_w[k]$
35	1200,178	1224,744	1249,7	49,522
36	1249,7	1275,055	1300,816	51,116
37	1300,816	1326,992	1353,592	52,776
38	1353,592	1380,623	1408,094	54,502
39	1408,094	1436,014	1464,392	56,298
40	1464,392	1493,237	1522,559	58,167
41	1522,559	1552,366	1582,668	60,109
42	1582,668	1613,474	1644,795	62,128
43	1644,795	1676,641	1709,021	64,226
44	1709,021	1741,946	1775,427	66,406
45	1775,427	1809,474	1844,098	68,671
46	1844,098	1879,31	1915,121	71,023
47	1915,121	1951,543	1988,587	73,466
48	1988,587	2026,266	2064,59	76,003
49	2064,59	2103,573	2143,227	78,637
50	2143,227	2183,564	2224,597	81,371
51	2224,597	2266,34	2308,806	84,208
52	2308,806	2352,008	2395,959	87,154
53	2395,959	2440,675	2486,169	90,21
54	2486,169	2532,456	2579,551	93,382
55	2579,551	2627,468	2676,223	96,672
56	2676,223	2725,832	2776,309	100,086
57	2776,309	2827,672	2879,937	103,627
58	2879,937	2933,12	2987,238	107,302
59	2987,238	3042,309	3098,35	111,112
60	3098,35	3155,379	3213,415	115,065
61	3213,415	3272,475	3332,579	119,164
62	3332,579	3393,745	3455,993	123,415
63	3455,993	3519,344	3583,817	127,823
64	3583,817	3649,432	3716,212	132,395
65	3716,212	3784,176	3853,348	137,136
66	3853,348	3923,748	3995,399	142,051
67	3995,399	4068,324	4142,547	147,148
68	4142,547	4218,09	4294,979	152,432
69	4294,979	4373,237	4452,89	157,911
70	4452,89	4533,963	4616,482	163,592
71	4616,482	4700,473	4785,962	169,48
72	4785,962	4872,978	4961,548	175,585
73	4961,548	5051,7	5143,463	181,915
74	5143,463	5236,866	5331,939	188,476
75	5331,939	5428,712	5527,217	195,278

ТАБЛИЦА 6 (окончание)

Группа	Нижняя частота/Гц	Центральная частота/Гц	Верхняя частота/Гц	Ширина полосы/Гц
k	$f_l[k]$	$f_c[k]$	$f_u[k]$	$f_w[k]$
76	5527,217	5627,484	5729,545	202,329
77	5729,545	5833,434	5939,183	209,637
78	5939,183	6046,825	6156,396	217,214
79	6156,396	6267,931	6381,463	225,067
80	6381,463	6497,031	6614,671	233,208
81	6614,671	6734,42	6856,316	241,646
82	6856,316	6980,399	7106,708	250,392
83	7106,708	7235,284	7366,166	259,458
84	7366,166	7499,397	7635,02	268,854
85	7635,02	7773,077	7913,614	278,594
86	7913,614	8056,673	8202,302	288,688
87	8202,302	8350,547	8501,454	299,152
88	8501,454	8655,072	8811,45	309,996
89	8811,45	8970,639	9132,688	321,237
90	9132,688	9297,648	9465,574	332,887
91	9465,574	9636,52	9810,536	344,962
92	9810,536	9987,683	10 168,013	357,477
93	10 168,013	10 351,586	10 538,46	370,447
94	10 538,46	10 728,695	10 922,351	383,891
95	10 922,351	11 119,49	11 320,175	397,824
96	11 320,175	11 524,47	11 732,438	412,264
97	11 732,438	11 944,149	12 159,67	427,231
98	12 159,67	12 379,066	12 602,412	442,742
99	12 602,412	12 829,775	13 061,229	458,817
100	13 061,229	13 296,85	13 536,71	475,48
101	13 536,71	13 780,887	14 029,458	492,748
102	14 029,458	14 282,503	14 540,103	510,645
103	14 540,103	14 802,338	15 069,295	529,192
104	15 069,295	15 341,057	15 617,71	548,415
105	15 617,71	15 899,345	16 186,049	568,339
106	16 186,049	16 477,914	16 775,035	588,986
107	16 775,035	17 077,504	17 385,42	610,385
108	17 385,42	17 690,045	18 000	614,58

ТАБЛИЦА 7

Полосы частот модели уха на основе БПФ, используемой в расширенной версии

Группа	Нижняя частота/Гц	Центральная частота/Гц	Верхняя частота/Гц	Ширина полосы/Гц
k	$f_l[k]$	$f_c[k]$	$f_u[k]$	$f_w[k]$
0	80	103,445	127,023	47,023
1	127,023	150,762	174,694	47,671
2	174,694	198,849	223,257	48,563
3	223,257	247,95	272,959	49,702
4	272,959	298,317	324,055	51,096
5	324,055	350,207	376,805	52,75
6	376,805	403,884	431,478	54,673
7	431,478	459,622	488,353	56,875
8	488,353	517,707	547,721	59,368
9	547,721	578,434	609,885	62,164
10	609,885	642,114	675,161	65,277
11	675,161	709,071	743,884	68,723
12	743,884	779,647	816,404	72,52
13	816,404	854,203	893,091	76,687
14	893,091	933,119	974,336	81,245
15	974,336	1016,797	1060,555	86,219
16	1060,555	1105,666	1152,187	91,632
17	1152,187	1200,178	1249,7	97,513
18	1249,7	1300,816	1353,592	103,892
19	1353,592	1408,094	1464,392	110,801
20	1464,392	1522,559	1582,668	118,275
21	1582,668	1644,795	1709,021	126,354
22	1709,021	1775,427	1844,098	135,077
23	1844,098	1915,121	1988,587	144,489
24	1988,587	2064,59	2143,227	154,64
25	2143,227	2224,597	2308,806	165,579
26	2308,806	2395,959	2486,169	177,364
27	2486,169	2579,551	2676,223	190,054
28	2676,223	2776,309	2879,937	203,713
29	2879,937	2987,238	3098,35	218,414
30	3098,35	3213,415	3332,579	234,229
31	3332,579	3455,993	3583,817	251,238
32	3583,817	3716,212	3853,348	269,531
33	3853,348	3995,399	4142,547	289,199
34	4142,547	4294,979	4452,89	310,343
35	4452,89	4616,482	4785,962	333,072
36	4785,962	4961,548	5143,463	357,5

ТАБЛИЦА 7 (окончание)

Группа	Нижняя частота/Гц	Центральная частота/Гц	Верхняя частота/Гц	Ширина полосы/Гц
k	$f_l[k]$	$f_c[k]$	$f_u[k]$	$f_w[k]$
37	5143,463	5331,939	5527,217	383,754
38	5527,217	5729,545	5939,183	411,966
39	5939,183	6156,396	6381,463	442,281
40	6381,463	6614,671	6856,316	474,853
41	6856,316	7106,708	7366,166	509,85
42	7366,166	7635,02	7913,614	547,448
43	7913,614	8202,302	8501,454	587,84
44	8501,454	8811,45	9132,688	631,233
45	9132,688	9465,574	9810,536	677,849
46	9810,536	10 168,013	10 538,46	727,924
47	10 538,46	10 922,351	11 320,175	781,715
48	11 320,175	11 732,438	12 159,67	839,495
49	12 159,67	12 602,412	13 061,229	901,56
50	13 061,229	13 536,71	14 029,458	968,229
51	14 029,458	14 540,103	15 069,295	1039,837
52	15 069,295	15 617,71	16 186,049	1116,754
53	16 186,049	16 775,035	17 385,42	1199,371
54	17 385,42	17 690,045	18 000	614,58

Преобразование частоты в высоту тона выполняется с помощью алгоритма, описанного в следующем подразделе, где $F_{sp}[k_f]$ – энергетическое представление *выходных данных БПФ, взвешенных по наружному уху*:

$$F_{sp}[k_f, n] = |F_e[k_f, n]|^2 \quad (11)$$

или энергетическое представление сигнала ошибки

$$F_{sp}[k_f, n] = |F_{noise}[k_f, n]|^2 \quad (12)$$

соответственно. Расчет сигнала ошибки см. в пункте 3.4.

Выходными данными этого этапа обработки являются значения энергии частотных групп $P_e[k, n]$.

2.1.5.1 Псевдокод

/* выходные данные */

$F_{sp}[\]$ Энергия входных сигналов

/* выходные данные */

$P_e[\]$: Энергия, преобразованная в высоту тона

/* промежуточные начения*/

i : Индекс частотной группы

k : \ Индекс строки БПФ

Z	Количество частотных групп: 109 – для базовой версии 55 – для расширенной версии
fl[]	: Нижняя частота частотной группы
fu[]	: Верхняя частота частотной группы
Fres	: Постоянная разрешения по частоте

```

Fres = 48000/2048;
for(i=0; i<Z; i++ )
{
  Pe[i]=0;
  for(k=0;k<1024;k++)
  { /* строка внутри частотной группы */
    if( (( k-0.5)*Fres >= fl[i]) && ((k+0.5)*Fres <= fu[i]))
    {
      Pe[i] += Fsp[k];
    }
    /* внутри частотной группы */
    else if( (( k-0.5)*Fres < fl[i]) && ((k+0.5)*Fres > fu[i]))
    {
      Pe[i] += Fsp[k]*(fu[i]-fl[i])/Fres;
    }
    /* левая граница */
    else if( ((k-0.5)*Fres < fl[i]) && ((k+0.5)*Fres > fl[i]))
    {
      Pe[i] += Fsp[k]*( k+0.5)*Fres - fl[i])/Fres;
    }
    /* правая граница
    else if( ((k-0.5)*Fres < fu[i]) && ((k+0.5)*Fres > fu[i]);
    {
      Pe[i] += Fsp[k]*(fu[i]- (k-0.5)*Fres)/Fres;
    }
    /* строка за пределами частотной группы */
    else
    {
      Pe[i] += 0;
    }
  }

  /* предельный результат */
  Pe[i]=max(Pe[i],0.000000000001);
}

```

2.1.6 Добавление внутреннего шума

К энергиям в каждой частотной группе добавляется частотно-зависимое смещение P_{Thres} :

$$P_{Thres}[k] = 10^{0,4-0,364 \cdot \left(\frac{f_c[k]}{\text{кГц}}\right)^{-0,8}}; \quad (13)$$

$$P_p[k, n] = P_e[k, n] + P_{Thres}[k]. \quad (14)$$

Выходные данные этого этапа обработки $P[k, n]$ называются *шаблонами высоты тона*.

2.1.7 Расширение спектра

Шаблоны высоты тона $P_p[k, n]$ размываются по частоте с использованием функции расширения, зависящей от уровня. Функция расширения спектра представляет собой двустороннюю экспоненциальную функцию. Наклон нижней части всегда равен 27 дБ/барк, а наклон верхней части зависит от частоты и энергии.

Наклон рассчитывается по формуле:

$$\frac{S_u[k, L[k, n]]}{\text{дБ/барк}} = -24 - \frac{230 \text{ Гц}}{f_c[k]} + 0,2 \cdot L[k, n] / \text{дБ}; \quad (15)$$

$$S_l[k, L[k, n]] = 27 \frac{\text{дБ}}{\text{барк}}, \quad (16)$$

где:

$$L[k, n] = 10 \cdot \log_{10}(P_p[k, n]).$$

Расширение спектра выполняется независимо для каждой группы частот k :

$$E_2[k, n] = \frac{1}{\text{Norm}_{SP}[k]} \left(\sum_{j=0}^{Z-1} E_{line}[j, k, n] \right)^{0,4}, \quad (17)$$

где E_{line} задано уравнением:

$$E_{line}[j, k, n] = \begin{cases} \frac{10^{\frac{L[j, n]}{10}} \cdot 10^{\frac{-res \cdot (j-k) \cdot S_l[j, L[j, n]]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot S_l[j, L[j, n]]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot S_u[j, L[j, n]]}{10}}} & \text{при } k < j; \\ \frac{10^{\frac{L[j, n]}{10}} \cdot 10^{\frac{res \cdot (k-j) \cdot S_u[j, L[j, n]]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot S_l[j, L[j, n]]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot S_u[j, L[j, n]]}{10}}} & \text{при } k \geq j. \end{cases} \quad (18)$$

$\text{Norm}_{SP}[k]$ вычисляется по формуле:

$$\text{Norm}_{SP}[k] = \left(\sum_{j=0}^{Z-1} \tilde{E}_{line}[j, k] \right)^{0,4}, \quad (19)$$

где:

$$\tilde{E}_{line}[j, k] = \begin{cases} \frac{10^{\frac{-res \cdot (j-k) \cdot S_l[j,0]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot S_l[j,0]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot S_u[\mu,0]}{10}}} & \text{при } k < j; \\ \frac{10^{\frac{res \cdot (k-j) \cdot S_u[j,0]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot S_l[j,0]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot S_u[\mu,0]}{10}}} & \text{при } k \geq j, \end{cases} \quad (20)$$

а res – разрешение шкалы высоты тона в барках (0,25 для базовой версии и 0,5 для расширенной версии).

Шаблоны, полученные на этом этапе обработки $E_2[k, n]$, позднее используются для расчета шаблонов модуляции и называются *неразмытыми шаблонами возбуждения*.

2.1.8 Расширение во временной области

Чтобы смоделировать прямое маскирование, значения энергии в каждой частотной группе размываются по времени фильтрами нижних частот первого порядка. Постоянные времени зависят от центральной частоты каждой группы (как указано в уравнении (10) и таблице 6) и рассчитываются по формуле

$$\tau = \tau_{\min} + \frac{100 \Gamma_{\Pi}}{f_c[k]} \cdot (\tau_{100} - \tau_{\min}) \quad \left| \begin{array}{l} \tau_{100} = 0,030 \text{ с;} \\ \tau_{\min} = 0,008 \text{ с.} \end{array} \right. \quad (21)$$

Фильтры нижних частот первого порядка рассчитываются по формуле

$$E_f[k, n] = a \cdot E_f[k, n-1] + (1-a) \cdot E_2[k, n]; \quad (22)$$

$$E[k, n] = \max(E_f(k, n), E_2(k, n)), \quad (23)$$

где a рассчитывается с использованием приведенных выше постоянных времени по формуле:

$$a = e^{-\frac{4}{187,5} \cdot \frac{1}{\tau}}; \quad (24)$$

n – фактический номер кадра, k – индекс группы и $E_f[k, 0] = 0$.

Шаблоны, полученные на этом этапе обработки $E[k, n]$, называются *шаблонами возбуждения*.

2.1.9 Порог маскирования

Маскирование описывает эффект, при котором более слабый, но отчетливо слышимый сигнал становится неслышимым при наличии более громкого сигнала. Этот порог рассчитывается путем взвешивания шаблонов возбуждения с помощью весовой функции $m[k]$:

$$m[k] = \begin{cases} 3,0 & \text{при } k \cdot res \leq 12; \\ 0,25 \cdot k \cdot res & \text{при } k \cdot res > 12; \end{cases} \quad (25)$$

$$M[k, n] = \frac{E[k, n]}{10^{\frac{m[k]}{10}}}. \quad (26)$$

Шаблоны, полученные на этом этапе обработки $M[k, n]$, называются *шаблонами маски*.

2.2 Модель уха на основе набора фильтров

2.2.1 Обзор

РИСУНОК 10

Модель периферического уха и предварительная обработка шаблонов возбуждения для модели, основанной на наборе фильтров



На входе модели уха на основе набора фильтров исследуемый и эталонный сигналы настраиваются до предполагаемого уровня воспроизведения и передаются через фильтр верхних частот для удаления постоянных и инфразвуковых составляющих сигналов. Затем сигналы разлагаются на полосовые сигналы с помощью линейных фазовых фильтров, равномерно распределенных по шкале воспринимаемой высоты тона. К полосовым сигналам применяется частотно-зависимое взвешивание для моделирования спектральных характеристик наружного и среднего уха. Зависимое от уровня спектральное разрешение слуховых фильтров моделируется сверткой выходных сигналов в частотной области с функцией расширения спектра, зависящей от уровня.

Огибающие сигналов рассчитываются путем применения преобразования Гильберта к полосовым сигналам (выпрямление), а для моделирования обратного маскирования применяется свертка во временной области с оконной функцией. Затем добавляется частотно-зависимое смещение, которое учитывает внутренний шум в слуховой системе и моделирует порог в тишине. Наконец, выполняется вторая свертка во временной области с использованием экспоненциальной функции расширения спектра, которая учитывает прямое маскирование.

Полученные *шаблоны возбуждения* используются для расчета *шаблонов удельной громкости*, а шаблоны, полученные до окончательного расширения во временной области (*неразмытые шаблоны возбуждения*), используются для расчета *шаблонов модуляции*. Вместе с самими шаблонами возбуждения они составляют основу для расчета выходных значений модели. Чтобы отделить влияние устойчивой частотной характеристики испытываемого устройства от других искажений, диаграммы возбуждения исследуемого и эталонного сигналов спектрально адаптируются друг к другу (адаптация). Шаблоны модуляции и шаблоны удельной громкости рассчитываются как по адаптированным, так и по неадаптированным шаблонам возбуждения.

2.2.2 Подвыборка

На выходе набора фильтров сигналы субдискретизируются с коэффициентом 32, а после первого расширения во временной области – с коэффициентом 6 (см. рисунок 11).

2.2.3 Настройка уровня воспроизведения

Коэффициент масштабирования входного сигнала рассчитывается на основе предполагаемого уровня воспроизведения полномасштабного входного сигнала по формуле

$$fac = \frac{10^{L_{\max}/20}}{32767}. \quad (27)$$

Если точный уровень воспроизведения неизвестен, рекомендуется установить L_{\max} равным 92 дБ_{SPL}.

2.2.4 Режекторный фильтр постоянного тока

Поскольку блок фильтров чувствителен к субзвуковым частотам во входных сигналах, к входным сигналам применяется режекторный фильтр постоянного тока. Используется фильтр верхних частот Баттерворта четвертого порядка с частотой среза 20 Гц. Фильтр реализован как каскад из двух БИХ-фильтров второго порядка:

$$y_n = x_n - 2x_{n-1} + x_{n-2} + b_1 y_{n-1} + b_2 y_{n-2} \quad (28)$$

с коэффициентами для первого блока:

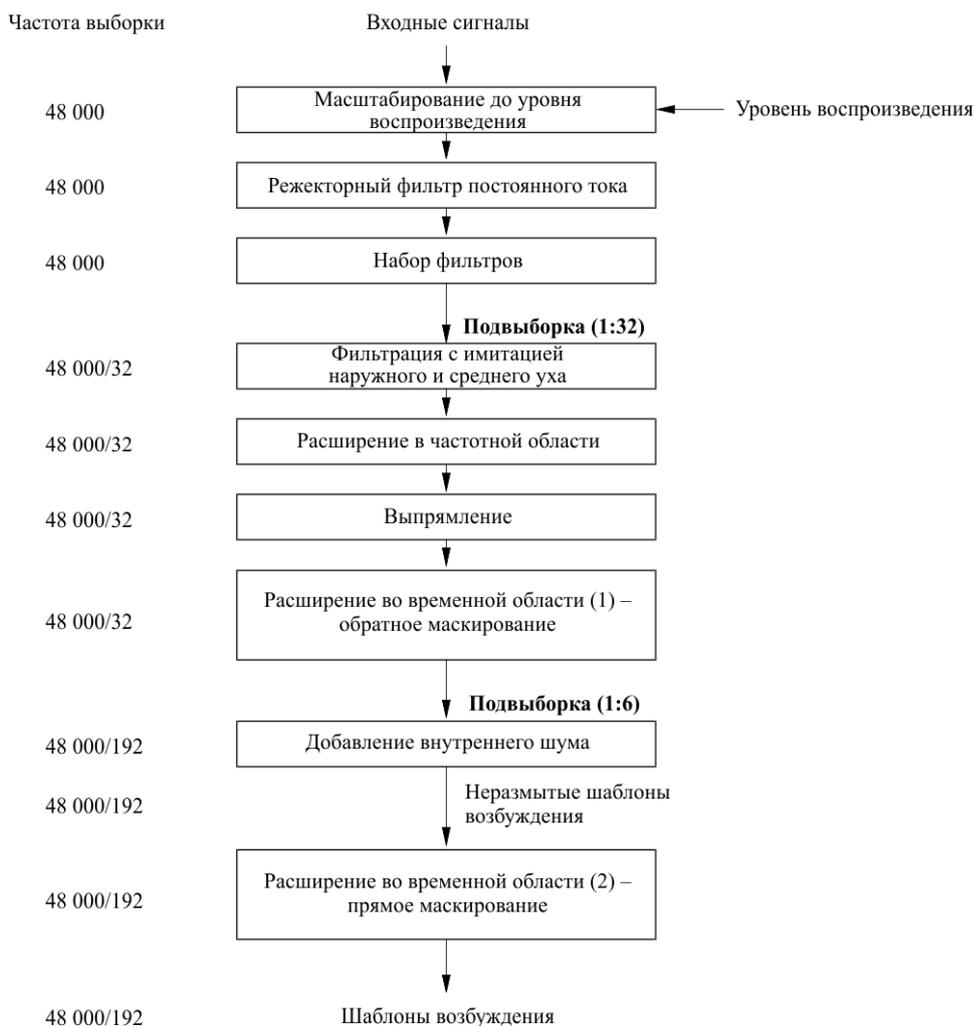
$$b_{1,2} = 1,99517, -0,995174$$

и коэффициентами для второго блока:

$$b_{1,2} = 1,99799, -0,997998.$$

РИСУНОК 11

Подвыборка в модели периферического уха на основе набора фильтров



BS.1387-11

2.2.5 Набор фильтров

Набор фильтров состоит из 40 пар фильтров для каждого канала исследуемого и эталонного сигналов. Фильтры расположены на равном расстоянии друг от друга и имеют постоянную абсолютную полосу пропускания по отношению к шкале слухового тона. Каждая пара фильтров состоит из двух фильтров с одинаковой частотной характеристикой, но разницей фазовых характеристик в 90° . Таким образом, выходной сигнал второго фильтра представляет собой результат преобразования Гильберта по отношению к выходному сигналу первого фильтра (или мнимую часть, если предполагается, что первый фильтр представляет действительную часть комплексного сигнала). Огибающие их импульсных характеристик имеют форму \cos^2 . Фильтры определяются таблицей 8 и уравнением (29) (где k – индекс фильтра, n – индекс временной выборки, а T – время между двумя выборками: $T = 1/48\,000$). Их можно реализовать как КИХ-фильтры, используя в качестве коэффициентов значения $h_{re}(k, n)$ и $h_{im}(k, n)$. Если входные сигналы ограничены по времени, то выходные данные фильтра также можно вычислить с помощью очень быстрых рекуррентных алгоритмов:

$$\begin{aligned}
 h_{re}(k, n) &= \frac{4}{N[k]} \cdot \sin^2 \left(\pi \cdot \frac{n}{N[k]} \right) \cdot \cos \left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2} \right) \cdot T \right) \\
 h_{im}(k, n) &= \frac{4}{N[k]} \cdot \sin^2 \left(\pi \cdot \frac{n}{N[k]} \right) \cdot \sin \left(2\pi \cdot f_c[k] \cdot \left(n - \frac{N[k]}{2} \right) \cdot T \right) \\
 h_{re}(k, n) &= h_{im}(k, n) = 0
 \end{aligned}
 \left. \begin{array}{l}
 0 \leq n < N[k]; \\
 n < 0; \\
 n \geq N[k].
 \end{array} \right\} \quad (29)$$

ТАБЛИЦА 8

Центральная частота, длина импульсной характеристики и дополнительная задержка для каждого фильтра

Индекс фильтра	Центральная частота/Гц	Длина импульсной характеристики/выборка	Дополнительная задержка/выборка
k	$f_c[k]$	$N[k]$	$D[k]$
0	50,00	1456	1
1	116,19	1438	10
2	183,57	1406	26
3	252,82	1362	48
4	324,64	1308	75
5	399,79	1244	107
6	479,01	1176	141
7	563,11	1104	177
8	652,97	1030	214
9	749,48	956	251
10	853,65	884	287
11	966,52	814	322
12	1089,25	748	355
13	1223,10	686	386
14	1369,43	626	416
15	1529,73	570	444
16	1705,64	520	469
17	1898,95	472	493
18	2111,64	430	514
19	2345,88	390	534
20	2604,05	354	552
21	2888,79	320	569
22	3203,01	290	584
23	3549,90	262	598
24	3933,02	238	610
25	4356,27	214	622
26	4823,97	194	632
27	5340,88	176	641

ТАБЛИЦА 8 (окончание)

Индекс фильтра	Центральная частота/Гц	Длина импульсной характеристики/выборка	Дополнительная задержка/выборка
k	$f_c[k]$	$M[k]$	$D[k]$
28	5912,30	158	650
29	6544,03	144	657
30	7242,54	130	664
31	8014,95	118	670
32	8869,13	106	676
33	9813,82	96	681
34	10 858,63	86	686
35	12 014,24	78	690
36	13 292,44	70	694
37	14 706,26	64	697
38	16 270,13	58	700
39	18 000,02	52	703

Центральные частоты находятся в диапазоне от 50 Гц до 18 000 Гц. Шкала слухового тона рассчитывается на основе приближения, предложенного в [Schroeder *et al*, 1979]:

$$z / \text{барк} = 7 \cdot \operatorname{arsinh} \left(\frac{f / \text{Гц}}{650} \right). \quad (30)$$

Единицей высоты тона служит *барк* (хотя эта шкала не совсем соответствует шкале Барка, определенной в [Zwicker and Feldtkeller, 1967]).

Чтобы задержки для всех фильтров были равными, входной сигнал каждого фильтра задерживается на D выборок, где D – половина разности между длиной его импульсной характеристики и длиной импульсной характеристики фильтра с самой длинной импульсной характеристикой³:

$$D[k] = 1 + \frac{1}{2} (N[0] - N[k]). \quad (31)$$

Выходные данные фильтра субдискретизируются с коэффициентом 32, то есть выходные значения для всех фильтров вычисляются для каждой 32-й входной выборки⁴.

³ Дополнительная задержка одной выборки для реализации не требуется. Отметим, что эталонная реализация, используемая для проверки соответствия, включает эту дополнительную задержку.

⁴ На самом деле огибающие фильтров в верхних полосах частот не обязательно удовлетворяют теореме дискретизации. Несмотря на то что наложение спектров может происходить только при очень специфических условиях (то есть при модулировании высокочастотных компонентов частотами выше 1,5 кГц) и проблемы, связанные с такими эффектами, никогда не возникали в известных базах данных, следует учитывать, что проблемы наложения спектров могут возникнуть, особенно при использовании искусственных исследуемых сигналов.

2.2.6 Фильтрация с имитацией наружного и среднего уха

Частотная характеристика наружного и среднего уха моделируется частотно-зависимой весовой функцией, которая применяется к выходным данным фильтра (уравнение (32)):

$$W[k]/\text{дБ} = -0,6 \cdot 3,64 \cdot \left(\frac{f_c[k]}{\text{кГц}}\right)^{-0,8} + 6,5 \cdot e^{-0,6 \cdot \left(\frac{f_c[k]}{\text{кГц}}\right)^{-3,3}} - 10^{-3} \cdot \left(\frac{f_c[k]}{\text{кГц}}\right)^{3,6}. \quad (32)$$

2.2.6.1 Псевдокод

/ входные данные */*

out_re,out_im : Выходные данные набора фильтров
(действительная и мнимая части)

W : Весовая функция (см. уравнение (32))

/ выходные данные */*

out_re,out_im : Выходные данные набора фильтров

/ промежуточные значения */*

k : Индекс фильтра

Wt : Весовой коэффициент

/ фильтрация с имитацией наружного и среднего уха */*

for(k = 0..39)

{

Wt = pow(10,W[k]/20)

out_re[k] *= Wt;

out_im[k] *= Wt;

}

2.2.7 Расширение в частотной области

Выходные значения набора фильтров размываются по частоте с использованием функции расширения спектра, зависящей от уровня. Функция расширения спектра представляет собой двустороннюю экспоненту. Наклон нижней части характеристики всегда составляет 31 дБ/барк, а наклон верхней варьируется от -24 до -4 дБ/барк.

Наклон верхней части $s[k]$ рассчитывается по формуле:

$$\frac{s[k]}{\text{дБ/барк}} = \min \left(-4, -24 - \frac{230 \text{ Гц}}{f_c[k]} + 0,2 \cdot L[k]/\text{дБ} \right). \quad (33)$$

Уровень $L[k]$ рассчитывается независимо для каждого канала фильтра путем возведения в квадрат абсолютного значения выходных данных фильтра и его преобразования в шкалу децибелов. Центральные частоты $f_c[k]$ взяты из таблицы 8. Линейные представления наклона сглаживаются по времени с помощью фильтра нижних частот первого порядка с постоянной времени 100 мс.

Расширение спектра осуществляется независимо для фильтров, представляющих действительную часть сигналов, и фильтров, представляющих мнимые части сигналов (уравнение (29)). Частотное рассеяние сначала выполняется для наклона верхней части характеристики (зависящего от уровня), а затем – для наклона нижней части с использованием алгоритма БИХ-фильтра первого порядка.

2.2.7.1 Псевдокод

/ входные данные */*

out_re,out_im : Выходные данные набора фильтров
(действительная и мнимая части)

z[] : Критические полосы центральных частот полос фильтра
в барках (согласно таблице 8 и уравнению (30))

```

/* выходные данные */
A_re,A_im      :      Выходные шаблоны
/* промежуточные значения */
j,k            :      Индекс фильтра
a,b            :      Коэффициенты сглаживания по времени
dist          :      Константа для расчета перекрестных помех
L[ ]          :      Уровень на каждом выходе фильтра
s[ ]          :      Локальный наклон для распространения вверх
d1,d2         :      Буферы
/* статика */
               (Сохраняются значения предыдущего кадра; в начале
               измерения значения инициализируются нулями)
cl, cu[ ]     :      Доля сигнала с расширяемым спектром

```

```

/* расширение вверх, зависящее от уровня */
dist      = pow(0.1,(z[39]-z[0])/(39*20));
/* (z[39]-z[0])/39 – расстояние в барках между двумя соседними полосами фильтра */
a         = exp(-32/(48000*0.1));
b         = 1 - a;
for(k=0..39)
{
    A_re[k]  = out_re[k];
    A_im[k]  = out_im[k];
}
for(k=0..39)
{
    /* расчет наклона, зависящего от уровня */
    L[k]     = 10*log10(out_re[k]* out_re[k] + out_im[k]* out_im[k]);
    s[k]     = max(4,(24 + 230/fcentre[k] - 0.2*L[k]));
    /* расчет расширяемой доли и ее сглаживание по времени*/
    cu[k]    = a*pow(dist,s[k])+b*cu[k];
    /* расширение спектра набора k */
    d1      = out_re[k]
    d2      = out_im[k]
    for(j=k+1..39)
    {
        d1   *= cu[k];
        d2   *= cu[k];
        A_re[j] += d1;
        A_im[j] += d2;
    }
}
/* расширение вниз */
cl        = pow(dist,31);
d1        = 0;
d2        = 0;

```

```

for(k=39..0)
{
    /* расширение спектра набора k */
    d1 = d1 * c1 + A_re[k];
    d2 = d2 * c1 + A_im[k];
    A_re[k] = d1;
    A_im[k] = d2;
}

```

2.2.8 Выпрямление

Энергия на выходах фильтра рассчитывается путем сложения квадратов значений фильтров, представляющих действительную часть сигнала, и значений, представляющих мнимую часть сигнала:

$$E_0[k, n] = A_{re}[k, n]^2 + A_{im}[k, n]^2. \quad (34)$$

Все последующие операции производятся над этими значениями энергии.

2.2.9 Размытие во временной области (1) – обратное маскирование

Для моделирования обратного маскирования значения энергии на выходах фильтра размываются во времени с помощью КИХ-фильтра с импульсной характеристикой в форме \cos^2 с 12 отводами (что соответствует характеристике фильтра из 384 выборок при частоте дискретизации входного сигнала набора фильтров). После размытия по времени выходные данные субдискретизируются с коэффициентом 6. Полученные значения умножаются на калибровочный коэффициент $cal_1 = 0,9761$, что дает соответствующие выходные уровни для данного уровня воспроизведения:

$$E_1[k, n] = \frac{0,9761}{6} \cdot \sum_{i=0}^{11} E_0[k, 6n-i] \cdot \cos^2\left(\pi \cdot \frac{(i-5)}{12}\right). \quad (35)$$

2.2.10 Добавление внутреннего шума

После первого расширения во временной области к значению энергии в каждом канале фильтра добавляется частотно-зависимое смещение E_{Thres} :

$$E_{Thres}[k] = 10^{0,4 \cdot 0,364 \cdot \left(\frac{f_c[k]}{\text{кГц}}\right)^{-0,8}}; \quad (36)$$

$$E_2[k, n] = E_1[k, n] + E_{Thres}[k, n]. \quad (37)$$

Шаблоны, полученные на этом этапе обработки $E_2[k, n]$, позднее используются для расчета шаблонов модуляции и называются *неразмытыми шаблонами возбуждения*.

2.2.11 Размытие во временной области (2) – прямое маскирование

Чтобы смоделировать прямое маскирование, энергия в каждом канале фильтра размывается по времени фильтрами нижних частот первого порядка. Постоянные времени зависят от центральной частоты каждого фильтра (как указано в таблице 5) и рассчитываются по формуле:

$$\tau = \tau_{\min} + \frac{100 \text{ Гц}}{f_c[k]} \cdot (\tau_{100} - \tau_{\min}) \quad \left| \begin{array}{l} \tau_{100} = 0,020\text{с}; \\ \tau_{\min} = 0,004\text{с}. \end{array} \right. \quad (38)$$

Фильтры нижних частот первого порядка рассчитываются по формуле:

$$E[k, n] = a \cdot E[k, n-1] + (1-a) \cdot E_2[k, n], \quad (39)$$

где a рассчитывается на основе приведенных выше постоянных времени по формуле:

$$a = e^{-\frac{192}{48\,000 \cdot \tau}}. \quad (40)$$

Шаблоны, полученные на этом этапе обработки $E[k, n]$, называются *шаблонами возбуждения*.

3 Предварительная обработка шаблонов возбуждения

Большинство вычислений, описанных в этом разделе, используются с моделью уха на основе набора фильтров, а также с моделью уха на основе БПФ. Поскольку коэффициент субдискретизации и количество полос частот в этих двух моделях уха различаются, константы, зависящие от этого фактора, описываются с помощью переменных, зависящих от модели уха, – $StepSize$ и Z . Для модели уха на основе БПФ значение $StepSize$ составляет 1024, а Z – либо 55 (расширенная версия), либо 109 (базовая версия). Для модели уха на основе набора фильтров значение $StepSize$ составляет 192, а Z – 40. Если не указано иное, то все переменные и рекурсивные фильтры инициализируются нулем.

3.1 Адаптация уровня и шаблона

Чтобы компенсировать разницу уровней и линейные искажения между исследуемым и эталонным сигналом, средние уровни исследуемого и эталонного сигналов адаптируются друг к другу.

На первом этапе значения энергии в каждом канале фильтра сглаживаются фильтрами нижних частот первого порядка. Постоянные времени зависят от центральных частот фильтров и выбираются по формуле

$$\tau = \tau_{\min} + \frac{100 \text{ Гц}}{f_c[k]} \cdot (\tau_{100} - \tau_{\min}) \quad \left| \begin{array}{l} \tau_{100} = 0,050 \text{ с;} \\ \tau_{\min} = 0,008 \text{ с.} \end{array} \right. \quad (41)$$

Фильтры нижних частот первого порядка рассчитываются по формулам:

$$P_{Ref}[k, n] = a \cdot P_{Ref}[k, n-1] + (1-a) \cdot E_{Ref}[k, n]; \quad (42)$$

$$P_{Test}[k, n] = a \cdot P_{Test}[k, n-1] + (1-a) \cdot E_{Test}[k, n]. \quad (43)$$

где E_{Test} и E_{Ref} – шаблоны возбуждения, которые необходимо адаптировать друг к другу, а a рассчитывается на основе постоянных времени по формуле

$$a = e^{-\frac{StepSize}{48\,000 \cdot \tau}}. \quad (44)$$

3.1.1 Адаптация уровня

По входным шаблонам фильтров нижних частот P_{Test} и P_{Ref} рассчитывается мгновенный поправочный коэффициент $LevCorr$ по следующей формуле:

$$LevCorr[n] = \left(\frac{\sum_{k=0}^{Z-1} \sqrt{P_{Test}[k, n] \cdot P_{Ref}[k, n]}}{\sum_{k=0}^{Z-1} P_{Test}[k, n]} \right)^2. \quad (45)$$

Если этот поправочный коэффициент больше единицы, то эталонный сигнал делится на поправочный коэффициент, в противном случае исследуемый сигнал умножается на поправочный коэффициент:

$$E_{L,Ref}[k, n] = E_{Ref}[k, n] / LevCorr[n] \quad | \quad LevCorr[n] > 1; \quad (46)$$

$$E_{L,Test}[k, n] = E_{Test}[k, n] \cdot LevCorr[n] \quad | \quad LevCorr[n] \leq 1. \quad (47)$$

3.1.2 Адаптация шаблона

Поправочные коэффициенты для каждого канала рассчитываются путем сравнения временных огибающих выходных данных фильтров исследуемого и эталонного сигналов:

$$R[k, n] = \frac{\sum_{i=0}^n a[k]^i \cdot E_{L,Test}[k, n-i] \cdot E_{L,Ref}[k, n-i]}{\sum_{i=0}^n a[k]^i \cdot E_{L,Ref}[k, n-i] \cdot E_{L,Ref}[k, n-i]}. \quad (48)$$

Значения a рассчитываются, как указано выше (уравнение (44)), на основе постоянных времени, заданных уравнением (41). Если $R[k, n]$ больше единицы, то поправочный коэффициент для исследуемого сигнала устанавливается равным $R[k, n]^{-1}$, а для эталонного сигнала – единице. В противном случае поправочный коэффициент для эталонного сигнала устанавливается равным $R[k, n]$, а для исследуемого сигнала – единице:

$$\begin{aligned} R_{Test}[k, n] &= \frac{1}{R[k, n]}, & R_{Ref}[k, n] &= 1 & \left| R[k, n] \geq 1; \right. \\ R_{Test}[k, n] &= 1, & R_{Ref}[k, n] &= R[k, n] & \left| R[k, n] < 1. \right. \end{aligned} \quad (49)$$

Если знаменатель в уравнении (48) равен нулю (и, следовательно, коэффициент $R[k, n]$ не определен), а числитель больше нуля, то $R_{Test}[k, n]$ устанавливается равным нулю, а $R_{Ref}[k, n]$ – единице. Если числитель в уравнении (48) также равен нулю, то копируются значения $R_{Test}[k, n]$ и $R_{Ref}[k, n]$ для нижележащей полосы частот. Если нижележащая полоса частот отсутствует (то есть $k = 0$), то коэффициенты $R_{Test}[k, n]$ и $R_{Ref}[k, n]$ устанавливаются равными единице.

Поправочные коэффициенты усредняются по M каналам фильтра и сглаживаются по времени (уравнение (50)) с использованием тех же постоянных времени, что указаны выше (уравнения (41)–(44)). Ширина частотного окна M для модели уха на основе набора фильтров равна 3. Для модели уха на основе БПФ она равна соответственно 4 (расширенная версия) или 8 (базовая версия):

$$\begin{aligned} PattCorr_{Test}[k, n] &= a \cdot PattCorr_{Test}[k, n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Test}[k+i, n]; \\ PattCorr_{Ref}[k, n] &= a \cdot PattCorr_{Ref}[k, n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Ref}[k+i, n]; \end{aligned} \quad (50)$$

$$\left| \begin{aligned} M_1 = M_2 = \frac{M-1}{2} & \quad | \quad M \text{ нечетное;} \\ M_1 = \frac{M}{2} - 1, \quad M_2 = \frac{M}{2} & \quad | \quad M \text{ четное.} \end{aligned} \right.$$

На границах шкалы частот, где частотное окно выходит за пределы диапазона полос фильтра, ширина частотного окна соответственно уменьшается:

$$M_1 = \min(M_1, k); \quad M_2 = \min(M_2, z - k - 1); \quad M = M_1 + M_2 + 1. \quad (51)$$

Шаблоны входных сигналов, адаптированные по уровню, взвешиваются с помощью соответствующих поправочных коэффициентов $PattCorr_{Test/Ref}[k, n]$ для получения спектрально адаптированных шаблонов:

$$E_{P,Ref}[k, n] = E_{L,Ref}[k, n] \cdot PattCorr_{Ref}[k, n]; \quad (52)$$

$$E_{P,Test}[k, n] = E_{L,Test}[k, n] \cdot PattCorr_{Test}[k, n]. \quad (53)$$

3.2 Модуляция

По неразмытым шаблонам возбуждения $E_2[k, n]$ вычисляется упрощенная громкость путем возведения возбуждения в степень 0,3. Это значение и абсолютное значение его производной по времени размываются по времени:

$$\bar{E}_{der}[k, n] = a \cdot \bar{E}_{der}[k, n-1] + (1-a) \cdot \frac{48\,000}{StepSize} \cdot |E_2[k, n]^{0,3} - E_2[k, n-1]^{0,3}|; \quad (54)$$

$$\bar{E}[k, n] = a \cdot \bar{E}[k, n-1] + (1-a) \cdot E_2[k, n]^{0,3}. \quad (55)$$

Значения a рассчитываются как для уравнения (44), на основе постоянных времени, определяемых формулой:

$$\tau = \tau_0 + \frac{100 \text{ Гц}}{f_c} \cdot (\tau_{100} - \tau_0) \left| \begin{array}{l} \tau_{100} = 0,050 \text{ с;} \\ \tau_0 = 0,008 \text{ с.} \end{array} \right. \quad (56)$$

По полученным значениям \bar{E}_{der} и \bar{E} вычисляется показатель модуляции огибающей на выходе каждого фильтра:

$$Mod[k, n] = \frac{\bar{E}_{der}[k, n]}{1 + \bar{E}[k, n]/0,3}. \quad (57)$$

Значения \bar{E} также используются позднее при расчете разности показателей модуляции.

3.3 Громкость

Конкретные характеристики громкости исследуемого и эталонного сигналов рассчитываются по формуле:

$$N[k, n] = const \cdot \left(\frac{1}{s[k]} \cdot \frac{E_{Thres}[k]}{10^4} \right)^{0,23} \cdot \left[\left(1 - s[k] + \frac{s[k] \cdot E[k, n]}{E_{Thres}[k]} \right)^{0,23} - 1 \right], \quad (58)$$

как указано в [Zwicker and Feldtkeller, 1967]. Общая громкость исследуемого и эталонного сигналов рассчитывается как сумма всех значений удельной громкости, превышающих ноль, по всем каналам фильтра:

$$N_{total}[n] = \frac{24}{Z} \cdot \sum_{k=0}^{Z-1} \max(N[k, n], 0). \quad (59)$$

Коэффициент масштабирования выбран равным $const = 1,07664$ для модели периферического уха на основе БПФ и $const = 1,26539$ для модели периферического уха на основе набора фильтров, чтобы обеспечить общую громкость в один сон для синусоидального тона с уровнем звукового давления 40 дБ_{SPL} на частоте 1 кГц. Индекс порога s и возбуждение на уровне порога E_{Thres} рассчитываются по формулам:

$$E_{Thres}[k] = 10^{0,364 \cdot \left(\frac{f}{1 \text{ кГц}} \right)^{-0,8}} \quad (60)$$

и

$$s[k] = 10^{\frac{1}{10} \left(-2 - 2,05 \cdot \text{atn} \left(\frac{f}{4 \text{ кГц}} \right) - 0,75 \cdot \text{atn} \left(\left(\frac{f}{1600 \text{ Гц}} \right)^2 \right) \right)} \quad (61)$$

соответственно.

ПРИМЕЧАНИЕ. – Из-за разных моделей периферийного уха рассчитанная здесь громкость не идентична громкости, определенной в ИСО 532 (Акустика. Методика расчета уровня громкости, 1975 год).

3.4 Расчет сигнала ошибки

Сигнал ошибки вычисляется только в модели на основе БПФ. Он рассчитывается в частотной области путем определения разности между спектрами амплитуд, отфильтрованными наружным и средним ухом, для эталонного и исследуемого сигналов (см. пункт 2.1.4):

$$F_{noise}[k_f, n] = \left| \left| F_{eref}[k_f, n] \right| - \left| F_{etest}[k_f, n] \right| \right|. \quad (62)$$

F_{noise} преобразуется в тональную область с использованием алгоритма, описанного в пункте 2.1.5.

Выходные данные этого алгоритма $P_{noise}[n, k]$ называются *шаблонами шума*.

4 Расчет выходных переменных модели

4.1 Обзор

ТАБЛИЦА 9

Обзор выходных переменных модели, используемых для прогнозирования базового качества звука

Выходная переменная модели (MOV)	Рассчитана для модели уха на основе		Используется в версии	
	БПФ	набора фильтров	базовой	расширенной
WinModDiff1 _B	Да	Нет	Да	Нет
AvgModDiff1 _B	Да	Нет	Да	Нет
AvgModDiff2 _B	Да	Нет	Да	Нет
RmsModDiff _A	Нет	Да	Нет	Да
RmsNoiseLoud _B	Да	Нет	Да	Нет
RmsNoiseLoudAsym _A	Нет	Да	Нет	Да
AvgLinDist _A	Нет	Да	Нет	Да
BandwidthRef _B	Да	Нет	Да	Нет
BandwidthTest _B	Да	Нет	Да	Нет
Total NMR _B	Да	Нет	Да	Нет
RelDistFrames _B	Да	Нет	Да	Нет
Segmental NMR _B	Да	Нет	Нет	Да
MFPD _B	Да	Нет	Да	Нет
ADB _B	Да	Нет	Да	Нет
EHS _B	Да	Нет	Да	Да

4.2 Разность показателей модуляции

Разность показателей модуляции для временных огибающих исследуемого и эталонного сигналов измеряется путем вычисления локальной разности показателей модуляции для каждого канала фильтра (уравнение (63)), где Mod_{test} и Mod_{Ref} получаются путем применения уравнения (57) к эталонному сигналу R_{test} :

$$ModDiff[k, n] = w \cdot \frac{|Mod_{test}[k, n] - Mod_{Ref}[k, n]|}{offset + Mod_{Ref}[k, n]};$$

$$\begin{cases} w = 1.0 & |Mod_{test}[k, n] > Mod_{Ref}[k, n], \\ w = negWt & |Mod_{test}[k, n] < Mod_{Ref}[k, n]. \end{cases} \quad (63)$$

Мгновенная разность показателей модуляции рассчитывается как среднее значение локальных разностей показателей модуляции по всем каналам фильтра (уравнение (64)):

$$ModDiff[n] = \frac{100}{Z} \sum_{k=0}^{Z-1} ModDiff[k, n]. \quad (64)$$

Порог в тишине учитывается зависящим от уровня весовым коэффициентом (уравнение (65)), рассчитанным на основе модифицированных шаблонов возбуждения для эталонного сигнала, как указано в уравнении (55), и функции внутреннего шума, как определено в уравнении (36) для модели уха на основе набора фильтров и уравнении (13) для модели уха на основе БПФ:

$$TempWt[n] = \sum_{k=0}^{Z-1} \frac{\bar{E}_{ref}[k, n]}{\bar{E}_{ref}[k, n] + levWt \cdot E_{Thres}[k]} ^{0,3}. \quad (65)$$

Временное усреднение мгновенных значений разности модуляции $ModDiff[n]$ с использованием весовых коэффициентов $TempWt[n]$ описано в пункте 5.2. Значения констант $negWt$, $offset$ и $levWt$ приведены в таблице 10.

ТАБЛИЦА 10

Выходные переменные модели, оценивающие общую разность показателей модуляции

MOV (Xxx=Win/Avg/Rms)	negWt	offset	levWt
$XxxModDiff1_B$	1	1	100
$XxxModDiff2_B$	0,1	0,01	100
$XxxModDiff_A$	1	1	1

4.2.1 RmsModDiff_A

Выходная переменная модели $RmsModDiff_A$ представляет собой среднеквадратичное значение разности показателей модуляции, рассчитанной для модели уха на основе набора фильтров. Сведения о временном усреднении см. в пункте 5.2.2, а значения констант приведены в таблице 10.

4.2.2 WinModDiff1_B

MOV $WinModDiff1_B$ – это скользящее среднее значение разности показателей модуляции, рассчитанное для модели уха на основе БПФ. Сведения о временном усреднении см. в пункте 5.2.3, а значения констант приведены в таблице 10. Временной весовой коэффициент, указанный в уравнении (65), для этой MOV не применяется.

4.2.3 AvgModDiff1_B и AvgModDiff2_B

MOV $AvgModDiff1_B$ и $AvgModDiff2_B$ представляют собой линейное среднее значение разности показателей модуляции, рассчитанное для модели уха на основе БПФ. Разница между $AvgModDiff2_B$ и $AvgModDiff1_B$ заключается в том, что константы выбираются по-разному. Сведения о временном усреднении см. в пункте 5.2.1, а значения констант приведены в таблице 10.

4.3 Громкость шума

Эти выходные переменные модели оценивают парциальную громкость аддитивных искажений в присутствии маскирующего эталонного сигнала. Формула парциальной громкости (уравнение (66)) позволяет определить удельную громкость шума в соответствии с [Zwicker and Feldtkeller, 1967] в отсутствие маскирующего сигнала и что-то вроде отношения сигнала шума к маскирующему сигналу, если шум очень мал по сравнению с ним.

Парциальная громкость шума рассчитывается по формуле:

$$NL[k, n] = \left(\frac{1}{s_{test}} \cdot \frac{E_{Thres}}{E_0} \right)^{0,23} \cdot \left[\left(1 + \frac{\max(s_{test} \cdot E_{test} - s_{ref} \cdot E_{ref}, 0)}{E_{Thres} + s_{ref} \cdot E_{ref} \cdot \beta} \right)^{0,23} - 1 \right], \quad (66)$$

где E_0 всегда равно 1, E_{Thres} – это функция внутреннего шума $E_{Thres}[k]$, определенная в уравнении (36), а s рассчитывается по формуле:

$$s = ThresFac_0 \cdot Mod[k, n] + S_0. \quad (67)$$

Если не указано иное, то в качестве входных данных используются *спектрально адаптированные шаблоны возбуждения* (см. пункт 3.1): $E_{Test} = E_{P, Test}[k, n]$ и $E_{Ref} = E_{P, Ref}[k, n]$. Коэффициент β , определяющий степень маскирования, рассчитывается по формуле:

$$\beta = \exp \left(-\alpha \cdot \frac{E_{test} - E_{ref}}{E_{ref}} \right). \quad (68)$$

Значения мгновенной громкости шума начинают учитываться только через 50 мс после того, как общая громкость левого или правого аудиоканалов превысит значение $N_{Thres} = 0,1$ сон как для исследуемого, так и для эталонного сигналов (см. пункт 5.2.4.2).

При усреднении спектра мгновенные значения нормируются по количеству полос фильтра на критическую полосу, а не по общему числу полос фильтра, то есть результат усреднения спектра умножается на 24.

Если мгновенная громкость шума ниже порогового значения NL_{min} , она устанавливается равной нулю.

ТАБЛИЦА 11

Выходные переменные модели, оценивающие общую громкость шума

MOV (Xxx=Win/Avg/Rms)	α	ThresFac ₀	S ₀	NL _{min}
XxxMissingComponents _B	1,5	0,15	1	0
XxxNoiseLoud _B	1,5	0,15	0,5	0
XxxMissingComponents _A	1,5	0,15	1	0
XxxNoiseLoud _A	2,5	0,3	1	0,1
XxxLinDist _A	1,5	0,15	1	0

4.3.1 RmsNoiseLoud_A

Выходная переменная модели $RmsNoiseLoud_A$ представляет собой среднеквадратичное значение громкости шума, рассчитанное для модели уха на основе набора фильтров. Сведения о временном усреднении см. в пункте 5.2.2, а значения констант приведены в таблице 11.

4.3.2 RmsMissingComponents_A

Выходная переменная модели *RmsMissingComponents_A* представляет собой среднеквадратичное значение громкости шума, рассчитанное для модели уха на основе набора фильтров. Она рассчитывается с использованием спектрально адаптированных шаблонов возбуждения исследуемого и эталонного сигналов, переставленных местами, чтобы получить громкость компонентов эталонного сигнала, которые теряются в исследуемом сигнале. Сведения о временном усреднении см. в пункте 5.2.2, а значения констант приведены в таблице 11.

4.3.3 RmsNoiseLoudAsym_A

Выходная переменная модели *RmsNoiseLoudAsym_A* представляет собой взвешенные среднеквадратичные значения громкости шума (см. пункт 4.3.1) и громкости потерянных компонентов сигнала (см. пункт 4.3.2), рассчитанных для модели уха на основе набора фильтров:

$$RmsNoiseLoudAsym = RmsNoiseLoud + 0,5 \cdot RmsMissingComponents \quad (69)$$

4.3.4 AvgLinDist_A

Выходная переменная модели *AvgLinDist_A* измеряет громкость компонентов сигнала, потерянных во время спектральной адаптации исследуемого и эталонного сигналов. В качестве эталонного сигнала используется спектрально адаптированное возбуждение эталонного сигнала, а в качестве исследуемого – неадаптированное возбуждение эталонного сигнала. Эта MOV рассчитывается для модели уха на основе набора фильтров. Сведения о временном усреднении см. в пункте 5.2.1, а значения констант приведены в таблице 11.

4.3.5 RmsNoiseLoud_B

Выходная переменная модели *RmsNoiseLoud_B* представляет собой среднеквадратичное значение громкости шума, рассчитанного для модели уха на основе БПФ. Сведения о временном усреднении см. в пункте 5.2.2, а значения констант приведены в таблице 11.

4.4 Ширина полосы

Эти выходные значения модели оценивают среднюю ширину полосы частот исследуемого и эталонного сигналов в линиях БПФ.

Для каждого кадра локальная ширина полосы $Bw_{Ref}[n]$ и $Bw_{Test}[n]$ рассчитывается в соответствии с приведенным ниже псевдокодом.

4.4.1 Псевдокод

```

/* входные данные */
FLevRef[], FlevelTest[]      :      Уровень выходных сигналов БПФ в дБ
/* выходные данные */
BwRef, BwTest                :      Выходные шаблоны
/* промежуточные значения */
k                             :      Индекс линий БПФ
ZeroThreshold                :      Порог ширины полосы

```

```

Zero Threshold = FLevelTst(921);
BwRef = BwTst = 0.0;
for(k=921;k<1024;k++)
{
  ZeroThreshold=max(ZeroThreshold,FLevelTst(k));
}
  for (k = 920; k>=0; k--)
  {
    if (FLevelRef[k] >= 10.0 + ZeroThreshold)
    {
      BwRef = k + 1;
      break;
    }
  }
  for (k = BwRef-1; k>=0; k--)
  {
    if(FLeveltest[k] >= 5.0+ZeroThreshold)
    {
      BwTest=k + 1;
      break;
    }
  }
}

```

4.4.2 BandwidthRef_B и BandwidthTest_B

BandwidthRef_B – это линейное среднее значение BwRef, а BandwidthTest_B – линейное среднее значение BwTest. При усреднении учитываются только кадры с BwRef > 346. Кадры с низкой энергией в начале и конце сигналов игнорируются (см. пункт 5.2.4.4). Сведения об усреднении по времени см. в пункте 5.2.1.

4.5 Отношение шума к маске

Следующие значения модели рассчитываются по значениям шума и маскирующего сигнала.

Локальное NMR текущего кадра n определяется по формуле:

$$NMR_{local}[n] = 10 * \log_{10} \frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k, n]}{M[k, n]} \quad (70)$$

4.5.1 Общее NMR_B

Выходная переменная модели *Общее NMR_B* представляет собой линейное среднее значение отношения шума к маске, рассчитанное по формуле:

$$NMR_{tot} = 10 * \log_{10} \frac{1}{N} \sum_n \left(\frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k, n]}{M[k, n]} \right). \quad (71)$$

Кадры с низкой энергией в начале и конце сигналов игнорируются (см. пункт 5.2.4.4).

4.5.2 NMR_B сегмента

MOV NMR_B сегмента (*Segmental NMR_B*) представляет собой линейное среднее значение локального NMR . Сведения об усреднении по времени см. в пункте 5.2.1.

Кадры с низкой энергией в начале и конце сигналов игнорируются (см. пункт 5.2.4.4).

4.6 Доля кадров с помехами

MOV *доля кадров с помехами* (Relative Disturbed Frames_B – RelDistFrames_B) представляет собой отношение числа кадров, для которых:

$$\max_{\forall k} \left(10 \cdot \log \left(\frac{P_{noise}[k,n]}{M[k,n]} \right) \right) \geq 1,5 \text{ дБ}, \quad k \in [0, Z-1]$$

к общему количеству кадров сигнала.

Кадры с низкой энергией в начале и конце сигналов игнорируются (см. пункт 5.2.4.4).

4.7 Вероятность обнаружения

MOV, определенные в этом разделе, основаны на значениях $\tilde{E}[k,n]$ (полоса k , кадр n), которые представляют собой *шаблоны возбуждения* $E[k, n]$, выраженные в децибелах:

$$\tilde{E}[k, n] = 10 \cdot \log_{10}(E[k, n]). \quad (72)$$

Для каждого кадра n :

выполняются следующие шаги независимо для каждого канала c (c имеет значения "левый" и "правый"). *Логарифмические шаблоны возбуждения* – это $\tilde{E}_{ref}[k,n]$ для эталонного сигнала и $\tilde{E}_{test}[k, n]$ для исследуемого сигнала.

Для каждой полосы k производятся следующие расчеты:

- Расчет асимметричного среднего возбуждения:

$$L[k, n] = 0,3 \cdot \max(\tilde{E}_{ref}[k, n], \tilde{E}_{test}[k, n]) + 0,7 \cdot \tilde{E}_{test}[k, n]. \quad (73)$$

- Расчет эффективного размера шага обнаружения s . Следующая формула дает приближение к едва заметной разнице уровней, измеренной согласно [Zwicker and Fastl, 1990]:

если $L[k, n] > 0$:

$$s[k, n] = 5,95072 \cdot ((6,39468)/L[k, n])^{1,71332} + 9,01033 \times 10^{-11} \cdot L[k, n]^4 + 5,05622 \times 10^{-6} \times L[k, n]^3 - 0,00102438 \cdot L[k, n]^2 + 0,0550197 \cdot L[k, n] - 0,198719,$$

иначе

$$s[k, n] = 1,0 \times 10^{30}. \quad (74)$$

- Расчет ошибки e со знаком:

$$e[k, n] = \tilde{E}_{ref}[k, n] - \tilde{E}_{test}[k, n]. \quad (75)$$

- Если $\tilde{E}_{ref}[k, n] > \tilde{E}_{test}[k, n]$, то крутизна наклона b устанавливается равной 4,0, в противном случае 6,0. Это моделирует эффект, заключающийся в том, что увеличение энергии исследуемого сигнала по сравнению с эталонным сигналом заметнее, чем уменьшение.

- Расчет масштабного коэффициента a :

$$a[k, n] = \frac{10^{\frac{\log_{10}(\log_{10}(2,0))}{b}}}{s[k, n]}. \quad (76)$$

- Расчет вероятности обнаружения. Уравнение (76) устанавливает масштабный коэффициент a таким, что если $e[k, n]$ равно $s[k, n]$, то $p_c[k, n]$ становится равным 0,5:

$$p_c[k, n] = 1 - 10^{(-a[k, n] \cdot e[k, n])^b}. \quad (77)$$

– Расчет общего количества шагов выше порога:

$$q_c[k, n] = \frac{|\text{INT}(e[k, n])|}{s[k, n]} . \quad (78)$$

– Вероятность бинаурального обнаружения равна:

$$p_{bin}[k, n] = \max(p_{left}[k, n], p_{right}[k, n]). \quad (79)$$

– Количество шагов выше порога бинаурального канала составляет

$$q_{bin}[k, n] = \max(q_{left}[k, n], q_{right}[k, n]). \quad (80)$$

Полная вероятность обнаружения канала c кадра n равна:

$$P_c[n] = 1 - \prod_{\forall k} (1 - p_c[k, n]) , \quad (81)$$

где c может иметь значение *левый*, *правый* или *bin*. Общее количество шагов выше порога для канала c кадра n равно:

$$Q_c[n] = \sum_{\forall k} q_c[k, n] . \quad (82)$$

4.7.1 Максимальная вероятность обнаружения после фильтрации (MFPD_B)

Производится расчет сглаженной версии вероятности обнаружения для каждого канала c :

$$\tilde{P}_c[n] = (1 - c_0) \cdot P_c[n] + c_0 \cdot \tilde{P}_c[n - 1], \quad (83)$$

где $P_c[-1] = 0$. Константа c_0 зависит от величины *StepSize*:

$$c_0 = 0,9 \text{StepSize}/1024 \quad (84)$$

c_0 понижает чувствительность к очень кратковременным искажениям.

Производится расчет максимальной вероятности обнаружения после фильтрации (Maximum Filtered Probability of Detection – MFPD):

$$PM_c[n] = \max(PM_c[n - 1] \cdot c_1, \tilde{P}_c[n]), \quad (85)$$

где $PM_c[-1] = 0$. Константа c_1 зависит от величины *StepSize*:

$$c_1 = 0,99 \text{StepSize}/1024; \quad (86)$$

c_1 моделирует эффект, заключающийся в том, что в результате забывания искажения в начале аудиофрагмента менее значительны, чем в конце. Отметим, что эта константа полезна для моделирования испытаний с прослушиванием, в которых испытуемым не разрешается выбирать более короткие части фрагмента. Для предлагаемой модели, откалиброванной с использованием результатов испытаний с прослушиванием в соответствии с Рекомендацией МСЭ-R BS.1116, значение c_1 должно быть равно 1,0.

MOV *MFPD* – это значение $PM_{bin}[n]$ для последнего кадра.

4.7.2 Средний искаженный блок⁵ (ADB_B)

Подсчитывается количество действительных кадров с вероятностью обнаружения бинаурального канала $P_{bin}[n]$ выше 0,5 ($n_{distorted}$).

Для всех действительных кадров рассчитывается общее количество шагов выше порога бинаурального канала $Q_{bin}[n]$:

⁵ Термин "блок" в этом контексте эквивалентен термину "кадр".

$$Q_{sum} = \sum_{\forall n} Q_{bin}[n].$$

Производится расчет искажения среднего искаженного блока ADB:

- если $n_{distorted}$ равно нулю, то $ADB = 0$ (искажений не слышно);
- если $n_{distorted} = 0$ и $Q_{sum} > 0$, то $ADB = \log_{10} ((Q_{sum}) / n_{distorted})$;
- если $n_{distorted} = 0$ и $Q_{sum} = 0$, то $ADB = -0,5$.

4.8 Гармоническая структура сигнала ошибки

Эталонный сигнал, содержащий сильные гармоники (например, бас-кларнет, клавишин), имеет спектр, характеризующийся рядом регулярно расположенных пиков, разделенных глубокими впадинами. При некоторых условиях сигнал ошибки может наследовать эту структуру. Например, шум, смешанный с таким сигналом, с большей вероятностью останется незамаскированным при низком уровне сигнала в спектральных впадинах. В этом случае результирующий спектр сигнала ошибки будет содержать структуру, аналогичную исходному спектру, но со сдвигом по частоте в соответствии с местоположением впадин. Эта структура может привести к искажению тональных качеств, что повысит заметность ошибки.

Сигнал ошибки определяется как разность логарифмических спектров эталонного и обработанного сигналов, каждый из которых взвешен согласно частотной характеристике наружного и среднего уха (см. пункт 2.1.4, уравнение (7)). Шаблон возбуждения, полученный из психоакустической модели, здесь не используется, поскольку нелинейное преобразование частоты в шкалу Барка приведет к искажению гармонической структуры.

4.8.1 EHSB

Величина гармонической структуры определяется путем выявления и измерения наибольшего пика в спектре автокорреляционной функции. Каждая корреляция рассчитывается как косинус угла между двумя векторами по следующей формуле, где \vec{F}_0 – вектор ошибки, а \vec{F}_t – тот же вектор с задержкой на определенную величину. Длина корреляции равна максимальной задержке (то есть 256 в примере ниже):

$$C = \frac{\vec{F}_0 \cdot \vec{F}_t}{|\vec{F}_0| \cdot |\vec{F}_t|}. \quad (87)$$

Максимальная задержка для получения автокорреляционной функции представляет собой наибольшую степень числа 2, которая меньше половины номера частотного компонента БПФ, соответствующего 18 кГц.

Например, при частоте дискретизации 48 кГц и размере окна БПФ 2048 выборки компонент БПФ, соответствующий 18 кГц, равен $(18/24) \times 1024 = 768$. Следовательно, максимальная задержка будет равна 384. Фактическое количество задержек будет равно 256, что является наибольшей степенью числа 2, меньшей 384. Первое значение корреляционной функции получается путем выравнивания $F_t[0]$ с $F_0[0]$, а последнее – путем выравнивания $F_t[0]$ с $F_0[255]$.

Результирующий вектор корреляций обрабатывается нормализованным окном Ханна, и после удаления постоянной составляющей путем вычитания среднего значения вычисляется спектр мощности с помощью БПФ. Максимальный пик спектра после первой впадины определяет доминирующую частоту автокорреляционной функции. Среднее значение этого максимума по кадрам, умноженное на 1000,0, представляет собой переменную гармонической структуры сигнала ошибки (EHS).

5 Усреднение

5.1 Спектральное усреднение

Если в описаниях MOV не указано иное (см. пункт 4), то при усреднении локальных значений по диапазонам частот используется следующий алгоритм.

5.1.1 Линейное среднее

Линейное среднее значение рассчитывается по формуле:

$$AvgS = \frac{1}{Z} \cdot \sum_{k=0}^{Z-1} S[k], \quad (88)$$

где S – имя выходной переменной модели, а Z – количество полос частот.

5.2 Усреднение по времени

Если в описаниях MOV (пункт 4) не указано иное, то при усреднении мгновенных значений по времени используются один или несколько из следующих алгоритмов. Временной весовой коэффициент (если он применяется) обозначается символом W , а количество полос частот – символом Z .

5.2.1 Линейное среднее

Линейное среднее значение (префикс Avg) рассчитывается по формуле:

$$AvgX = \frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n], \quad (89)$$

где X означает имя MOV, а N – количество выборок по времени, для которых рассчитаны мгновенные значения X .

В качестве альтернативы при использовании взвешивания по времени (см. пункт 4.2) линейное среднее значение рассчитывается по формуле:

$$AvgX = \frac{\sum_{n=0}^{N-1} W[n] \cdot X[n]}{\sum_{n=0}^{N-1} W[n]}. \quad (90)$$

5.2.2 Среднеквадратичное значение

Среднеквадратичное значение (префикс Rms) рассчитывается по формуле:

$$RmsX = \sqrt{\frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n]^2}, \quad (91)$$

где X означает имя MOV, а N – количество выборок по времени, для которых рассчитаны мгновенные значения X .

В качестве альтернативы, при использовании взвешивания по времени (см. пункт 4.2) среднеквадратичное значение рассчитывается по формуле:

$$RmsX = \sqrt{Z} \cdot \sqrt{\frac{\sum_{n=0}^{N-1} W[n]^2 \cdot X[n]^2}{\sum_{n=0}^{N-1} W[n]^2}}. \quad (92)$$

5.2.3 Скользящее среднее

Скользящее среднее значение (префикс *Win*) рассчитывается по формуле:

$$WinX = \sqrt{\frac{1}{N-L+1} \cdot \sum_{n=L-1}^{N-1} \left(\frac{1}{L} \cdot \sum_{i=0}^{L-1} \sqrt{X[n-i]} \right)^4}, \quad (93)$$

где X означает имя MOV, N – количество выборок по времени, для которых рассчитаны мгновенные значения X , а L – длина скользящего временного окна с выборками по времени. Длина окна составляет приблизительно 100 мс, то есть $L = 4$ для модели уха на основе БПФ и 25 для модели уха на основе набора фильтров.

5.2.4 Выбор кадра

5.2.4.1 Усреднение с задержкой

Для MOV, использующих этот критерий, значения, рассчитанные в течение первой 0,5 с измерения, при усреднении по времени не учитываются. *Усреднение с задержкой* используется для следующих выходных переменных модели:

WinModDiff1, AvgModDiff1, AvgModDiff2, RmsNoiseLoudness, RmsNoiseLoudAsym,
RmsModDiff, AvgLinDist.

5.2.4.2 Порог громкости

Для MOV, использующих этот критерий, все мгновенные значения, рассчитанные до тех пор, пока не пройдут 50 мс с того момента, когда общая громкость одного из соответствующих аудиоканалов достигла значения N_{Thres} сонов как для исследуемого, так и для эталонного сигнала, при усреднении по времени не учитываются. *Порог громкости* используется только для MOV, описанных в пункте 4.3.

5.2.4.3 Энергетический порог

Если энергия последней половины кадра из 2048 выборок меньше 8000* в моноканале или в левом и правом каналах как эталонных, так и исследуемых данных, то кадр игнорируется. Кадры перекрываются на 50% и оценивается только полукадр, содержащий новые данные. Применение этого критерия предотвращает обработку кадров с очень малой энергией.

Этот критерий используется только для MOV, описанных в пункте 4.8.

5.2.4.4 Граница данных

Если обработанный файл содержит шум до или после допустимых данных эталонного файла, то относительная ошибка может быть очень большой, поскольку эталонный уровень приближается к $-\infty$. Если эта ошибка считается артефактом, то ее можно игнорировать, применив критерий отсечки пограничных данных.

* Это число относится к входным данным в формате 16-битового целого числа со знаком в диапазоне от $-32\,768$ до $32\,767$, используемом для компакт-дисков.

При первом открытии файлов определяются местоположения начала и конца фактических данных в эталонном файле. Начало или конец данных определяется как первая позиция от начала или конца файла, где сумма абсолютных значений пяти последовательных выборок в одном из соответствующих аудиоканалов превышает 200. Кадры, полностью выходящие за пределы этого диапазона, впоследствии игнорируются.

Этот критерий используется для MOV.

5.3 Усреднение по аудиоканалам

Если не указано иное, то в случае стереосигналов MOV левого и правого каналов линейно усредняются после усреднения по времени.

6 Оценка воспринимаемого базового качества звука

Воспринимаемое базовое качество звука оценивается путем сопоставления нескольких MOV с одним числом с использованием структуры искусственной нейронной сети с одним скрытым слоем.

6.1 Искусственная нейронная сеть

Функция активации нейронной сети представляет собой асимметричную сигмоиду:

$$\text{sig}(x) = \frac{1}{1 + e^{-x}}. \quad (94)$$

Сеть имеет входы I и узлы J в скрытом слое. Преобразование определяется набором входных коэффициентов масштабирования $a_{\min}[i]$, $a_{\max}[i]$, набором входных весовых коэффициентов $w_x[i]$, набором выходных весовых коэффициентов $w_y[j]$ и парой выходных коэффициентов масштабирования b_{\min} и b_{\max} . Входные данные преобразуются в *индекс искажения*:

$$DI = w_y[J] + \sum_{j=0}^{J-1} \left(w_y[j] \cdot \text{sig} \left(w_x[I, j] + \sum_{i=0}^{I-1} w_x[i, j] \cdot \frac{x[i] - a_{\min}[i]}{a_{\max}[i] - a_{\min}[i]} \right) \right), \quad (95)$$

который напрямую связан с оценкой *воспринимаемого базового качества звука* в виде *объективной оценки разницы* (ODG). Связь между *индексом искажения* и *объективной оценкой разницы* определяется следующим образом:

$$ODG = b_{\min} + (b_{\max} - b_{\min}) \cdot \text{sig}(DI). \quad (96)$$

6.2 Базовая версия

В базовой версии используется только модель уха на основе БПФ. В ней используются следующие MOV: *BandwidthRefB*, *BandwidthTestB*, *Total NMRB*, *WinModDiff1B*, *ADB*, *EHS*, *AvgModDiff1B*, *AvgModDiff2B*, *RmsNoiseLoudB*, *MFPD* и *RelDistFramesB*. Эти 11 MOV сопоставляются с одним индексом качества с использованием нейронной сети, как описано в пункте 6.1 (*Искусственная нейронная сеть*), с тремя узлами в скрытом слое. Параметры преобразования приведены в таблицах 12–16.

ТАБЛИЦА 12
MOV, используемые в базовой версии

MOV	Назначение
WinModDiff1 _B	Изменения модуляции (связанные с резкостью)
AvgModDiff1 _B	
AvgModDiff2 _B	
RmsNoiseLoud _B	Громкость искажений
BandwidthRef _B	Линейные искажения (частотная характеристика и т. д.)
BandwidthTest _B	
RelDistFrames _B	Частота слышимых искажений
Total NMR _B	Отношение шума к маске
MFPD _B	Вероятность обнаружения
ADB _B	
EHS _B	Гармоническая структура сигнала ошибки

ТАБЛИЦА 13
Входные коэффициенты масштабирования в базовой версии

Индекс (i)	MOV (x[i])	a _{min} [i]	a _{max} [i]
0	BandwidthRef _B	393,916656	921
1	BandwidthTest _B	361,965332	881,131226
2	Total NMR _B	-24,045116	16,212030
3	WinModDiff1 _B	1,110661	107,137772
4	ADB _B	-0,206623	2,886017
5	EHS _B	0,074318	13,933351
6	AvgModDiff1 _B	1,113683	63,257874
7	AvgModDiff2 _B	0,950345	1145,018555
8	RmsNoiseLoud _B	0,029985	14,819740
9	MFPD _B	0,000101	1
10	RelDistFrames _B	0	1

ТАБЛИЦА 14

Весовые коэффициенты для входных узлов в базовой версии

Индекс (i)	MOV ($x[i]$)	Узел 1 ($w_x[i,0]$)	Узел 2 ($w_x[i,1]$)	Узел 3 ($w_x[i,2]$)
0	BandwidthRef _B	-0,502657	0,436333	1,219602
1	BandwidthTest _B	4,307481	3,246017	1,123743
2	Total NMR _B	4,984241	-2,211189	-0,192096
3	WinModDiff1 _B	0,051056	-1,762424	4,331315
4	ADB _B	2,321580	1,789971	-0,754560
5	EHS _B	-5,303901	-3,452257	-10,814982
6	AvgModDiff1 _B	2,730991	-6,111805	1,519223
7	AvgModDiff2 _B	0,624950	-1,331523	-5,955151
8	RmsNoiseLoud _B	3,102889	0,871260	-5,922878
9	MFPD _B	-1,051468	-0,939882	-0,142913
10	RelDistFrames _B	-1,804679	-0,503610	-0,620456
11	bias	-2,518254	0,654841	-2,207228

ТАБЛИЦА 15

Весовые коэффициенты для выходных узлов в базовой версии

Узел 1 ($w_y[0]$)	Узел 2 ($w_y[1]$)	Узел 3 ($w_y[2]$)	Смещение ($w_y[3]$)
-3,817048	4,107138	4,629582	-0,307594

ТАБЛИЦА 16

Выходные коэффициенты масштабирования в базовой версии

	b_{\min}	b_{\max}
ODG	-3,98	0,22

6.3 Расширенная версия

В расширенной версии используются как модель уха на основе набора фильтров, так и модель на основе БПФ. Для нее используются MOV RmsModDiff_A, RmsNoiseLoudAsym_A, AvgLinDist_A, Segmental NMR_B и EHS_B. Эти пять MOV сопоставляются с одним индексом качества с использованием нейронной сети с пятью узлами в скрытом слое, как описано в пункте 6.1. Параметры преобразования приведены в таблицах 17–21.

ТАБЛИЦА 17

MOV, используемые в расширенной версии

MOV	Назначение
RmsNoiseLoudAsym _A	Громкость искажений
RmsModDiff _A	Изменения модуляции (связанные с резкостью)
AvgLinDist _A	Линейные искажения (частотная характеристика и т. д.)
Segmental NMR _B	Отношение шума к маске
EHS _B	Гармоническая структура сигнала ошибки

ТАБЛИЦА 18

Коэффициенты масштабирования для входных узлов в расширенной версии

Индекс (i)	MOV (x[i])	a _{min} [i]	a _{max} [i]
0	RmsModDiff _A	13,298751	2166,5
1	RmsNoiseLoudAsym _A	0,041073	13,24326
2	Segmental NMR _B	-25,018791	13,46708
3	EHS _B	0,061560	10,226771
4	AvgLinDist _A	0,024523	14,224874

ТАБЛИЦА 19

Входные весовые коэффициенты в расширенной версии

Индекс (i)	MOV (x[i])	Узел 1 (w _x [i,0])	Узел 2 (w _x [i,1])	Узел 3 (w _x [i,2])	Узел 4 (w _x [i,3])	Узел 5 (w ₄ [i,4])
0	RmsModDiff _A	21,211773	-39,913052	-1,382553	-14,545348	-0,320899
1	RmsNoiseLoudAsym _A	-8,981803	19,956049	0,935389	-1,686586	-3,238586
2	Segmental NMR _B	1,633830	-2,877505	-7,442935	5,606502	-1,783120
3	EHS _B	6,103821	19,587435	-0,240284	1,088213	-0,511314
4	AvgLinDist _A	11,556344	3,892028	9,720441	-3,287205	-11,031250
5	bias	1,330890	2,686103	2,096598	-1,327851	3,087055

ТАБЛИЦА 20

Весовые коэффициенты для выходных узлов в расширенной версии

Узел 1 (w _x [i,0])	Узел 2 (w _x [i,1])	Узел 3 (w _x [i,2])	Узел 4 (w _x [i,3])	Узел 5 (w ₄ [i,4])	Смещение (w _y [5])
-4,696996	-3,289959	7,004782	6,651897	4,009144	-1,360308

ТАБЛИЦА 21

Выходные коэффициенты масштабирования в расширенной версии

	b_{\min}	b_{\max}
ODG	-3,98	0,22

7 Соответствие реализаций**7.1 Общие положения**

В этом разделе представлен набор исследуемых элементов для проверки правильности реализации методики.

7.2 Выбор

Исследуемые элементы выбирались из базы данных 3 (DB3), которая использовалась для проверки моделей. Для упрощения испытаний из 84 элементов DB3 было выбрано подмножество, состоящее из 16 элементов. Основным критерием выбора было то, чтобы полученные значения MOV и DI (*индекса искажения*) охватывали широкий диапазон значений.

7.3 Настройки для испытания на соответствие

Исследуемые элементы доступны для загрузки в виде WAV-файлов (формат Microsoft RIFF). Все элементы дискретизировались с частотой 48 кГц в 16-битовом формате ИКМ. Эталонный и исследуемый сигналы, предоставленные МСЭ, уже адаптированы друг к другу по времени и уровню, поэтому никакой дополнительной компенсации усиления или задержки не требуется. Алгоритм измерения должен быть настроен на уровень прослушивания 92 дБ SPL.

7.4 Приемлемый интервал допуска

Чтобы соответствовать Рекомендации, рассчитанные значения DI должны воспроизводить значения, приведенные в таблицах 22 и 23, с допуском менее $\pm 0,02$ для всех исследуемых элементов⁶. Если реализация не дает результатов в пределах этого допуска, она не соответствует настоящей Рекомендации.

⁶ Для достижения этой точности следует использовать стандарт арифметики с плавающей запятой IEEE.

7.5 Исследуемые элементы

В следующих таблицах указаны названия эталонных и исследуемых элементов⁷, а также полученные значения DI. Таблица 22 относится к базовой версии, а таблица 23 – к расширенной.

ТАБЛИЦА 22

Исследуемые элементы и результирующие значения DI для базовой версии

Элемент	DI	ODG
acodsna.wav	1,304	-0,676
bcodtri.wav	1,949	-0,304
ccodsax.wav	0,048	-1,829
ecodsmg.wav	1,731	-0,412
fcodsb1.wav	0,677	-1,195
fcodtr1.wav	1,419	-0,598
fcodtr2.wav	-0,045	-1,927
fcodtr3.wav	-0,715	-2,601
gcodcla.wav	1,781	-0,386
icodsna.wav	-3,029	-3,786
kcodsme.wav	3,093	0,038
lcodhrp.wav	1,041	-0,876
lcodpip.wav	1,973	-0,293
mcodcla.wav	-0,436	-2,331
ncodsfe.wav	3,135	0,045
scodclv.wav	1,689	-0,435

ТАБЛИЦА 23

Исследуемые элементы и результирующие значения DI для расширенной версии

Элемент	DI	ODG
acodsna.wav	1,632	-0,467
bcodtri.wav	2,000	-0,281
ccodsax.wav	0,567	-1,300
ecodsmg.wav	1,594	-0,489
fcodsb1.wav	1,039	-0,877
fcodtr1.wav	1,555	-0,512
fcodtr2.wav	0,162	-1,711
fcodtr3.wav	-0,783	-2,662
gcodcla.wav	1,457	-0,573
icodsna.wav	-2,510	-3,664

⁷ Имена соответствующих эталонных элементов получаются путем замены подстроки "cod" в именах исследуемых элементов на "ref", например эталонным элементом для "bcodtri.wav" служит "breftri.wav".

ТАБЛИЦА 23 (окончание)

Элемент	DI	ODG
kcodsme.wav	2,765	-0,029
lcodhrp.wav	1,538	-0,523
lcodpip.wav	2,149	-0,219
mcodcla.wav	0,430	-1,435
ncodsfe.wav	3,163	0,050
scodclv.wav	1,972	-0,293

Прилагаемый документ 1 к Приложению 2

Процесс проверки

1 Общие сведения

В 1994 году МСЭ-R принял Вопрос МСЭ-R 210/10 – Методы объективной оценки воспринимаемого качества [звука], и была создана Целевая группа. Одним из первых действий было объявление открытого конкурса предложений, на который были представлены шесть моделей.

Были приложены значительные усилия для определения процедур процесса проверки. Было сочтено полезным составить первую базу данных, названную DB1, состоящую из материалов уже проведенных испытаний с прослушиванием. Основное внимание уделялось среднему и высокому качеству звука, поэтому учитывались только результаты испытаний с прослушиванием в соответствии с Рекомендацией МСЭ-R BS.1116. Материал этих испытаний представлял собой критический радиовещательный материал для кодеков с низкой битовой скоростью, таких как MPEG1 уровень II, MPEG1 уровень III, Dolby AC2, Mini Disc, NICAM и др. База данных 1 (DB1) была предназначена для предоставления разработчикам моделей общей платформы, состоящей из материалов, охватывающих широкий спектр искажений, различные кодеки и нарушения сигналов, вызванные цепочкой кодеков. Подробное описание испытаний, собранных в базе данных 1, приведено в Прилагаемом документе 2 к Приложению 2.

Очевидно, что объективный метод измерения воспринимаемого качества звука, имитирующий поведение человека, можно проверить только по базе данных, содержащей результаты субъективных испытаний. Для надлежащей проверки требуется база данных, основанная на незнакомом материале. По этой причине пришлось провести новые испытания с прослушиванием. Поскольку в идеале метод измерения должен быть ориентирован на любые типы артефактов, которые могут возникать в приложениях радиовещания, следует включать не только артефакты кодирования, но и более традиционные артефакты, такие как искажения и шум. Для удовлетворения этих требований в 1996 и 1997 годах были составлены соответственно база данных 2 и база данных 3. К кодекам, уже включенным в DB1, были добавлены Dolby AC-3 и AAC. Более подробная информация содержится в Прилагаемом документе 2 к Приложению 2.

При проверке следует учитывать неопределенности, присущие субъективным испытаниям с прослушиванием, часто представленные в виде доверительных интервалов. Размер доверительного интервала зависит от ряда факторов. Наиболее важными являются опыт испытуемых, процедуры обучения и контекст, в котором представлены исследуемые элементы, а также количество испытуемых.

Адаптация и проверка объективного метода, приведенного в настоящей Рекомендации, рассчитана на "слушателя со средним опытом". Для характеристики "слушателя со средним опытом" используются средние значения субъективных оценок качества с 95%-ми доверительными интервалами.

Субъективные испытания с прослушиванием очень чувствительны к различным факторам, влияющим на результаты. SDG как для базы данных 2, так и для базы данных 3 были созданы на трех разных опытных площадках, и в ряде исследований изучалась возможность объединения этих данных. Хотя не все исследования привели к одинаковым выводам, было сочтено целесообразным объединить данные, и эта объединенная база данных легла в основу проверки.

Процесс проверки был разделен на три этапа:

- этап 1: конкуренция;
- этап 2: сотрудничество;
- этап 3: окончательный выбор.

Эти этапы подробно описаны в следующих разделах.

2 Этап конкуренции

Были предложены шесть методов объективного измерения воспринимаемого качества звука (DIX, NMR, PAQM, PERCEVAL, POM, TTA), и было решено сравнить их эффективность с использованием базы данных 2 и подмножества базы данных 1. База данных 2 была создана в начале 1996 года. Отбор окончательного материала для испытаний стал результатом совместных усилий компаний SR (Швеция) и BBC (Соединенное Королевство). Испытания с прослушиванием проводились в NRK в Норвегии, DR в Дании и NHK в Японии. Статистический анализ данных испытаний подготовили компании Deutsche Telekom (Германия) и Terascom (Швеция). Объективные данные на этапе 1 были получены на нейтральной площадке (Swisscom, Швейцария). Затем разработчики модели получили первую половину базы данных 2 для окончательной адаптации методов (этап 2). Наконец, в Swisscom были получены новые объективные оценки разницы.

Анализ эффективности методов проводился компанией Terascom (Швеция), а также самими разработчиками. Хотя результаты некоторых из предложенных методов показали высокую корреляцию с SDG, был достигнут консенсус в отношении того, что ни один из методов не отвечает требованиям пользователей. Отдельное исследование показало, что ни один из предложенных методов не оказался существенно лучше остальных. Поэтому было решено совместными усилиями всех разработчиков создать усовершенствованный метод измерения. Эффективность нового метода нужно сравнивать с одним из уже существующих методов под названием модель В3.

3 Этап сотрудничества

Этап сотрудничества был основан на идее объединения лучших элементов различных методов в единый новый метод. Чтобы в максимальной степени удовлетворить потребности пользователей, было решено разработать две версии метода – версию, подходящую для реализаций в режиме реального времени, и версию, которой для достижения более высокой точности может потребоваться повышенная вычислительная мощность.

Была разработана процедура проверки новых методов, аналогичная процедуре для этапа конкуренции. Пришлось создать новую базу данных (DB3). Пункты и условия были окончательно определены весной 1997 года и обобщены в SR, Swisscom и BBC. Полное описание базы данных содержится в Прилагаемом документе 2 к Приложению 2. Субъективные испытания с прослушиванием проводились на трех опытных площадках: Deutsche Telekom, NHK и SR. На всех площадках применялся "двойной слепой метод с тремя входными сигналами и скрытым эталонным сигналом", описанный в Рекомендации МСЭ-R BS.1116. Результаты прослушивания были собраны в Швеции. Обширный статистический анализ результатов испытаний с прослушиванием был проведен в компании Terascom и других организациях. По итогам этого анализа некоторые слушатели были исключены из дальнейшего процесса оценки. Результаты с опытных площадок были объединены в базу данных 3.

Осенью 1997 года разработчикам было передано 52 элемента базы данных. Новые методы были адаптированы к новым данным. Поскольку несколько наборов параметров дали схожие результаты, решение об окончательном выборе по возможности откладывалось. Наконец, в Швейцарии для проверки новых методов на "неизвестном" наборе данных были использованы оставшиеся 32 элемента.

Кроме того, для проверки новых методов на "неизвестном" материале были использованы результаты нового испытания с прослушиванием, проведенного CRC (Канада). Процесс выбора и проверки пригодности описан в следующих разделах.

4 Проверка пригодности

Были проведены многочисленные испытания 18 указанных вариантов объективного метода измерений. В этом разделе описаны критерии выбора, а также результаты сравнения SDG с результатами, полученными с использованием 18 вариантов представленного метода измерения. Цель заключалась в выборе и проверке оптимальных версий, которые будут рекомендованы МСЭ.

Критерии выбора

Наиболее очевидным критерием проверки объективного метода служит корреляция между субъективными и объективными результатами. Кроме того, в целях проверки были введены еще два критерия, учитывающие надежность средних значений, – AES и схема допуска.

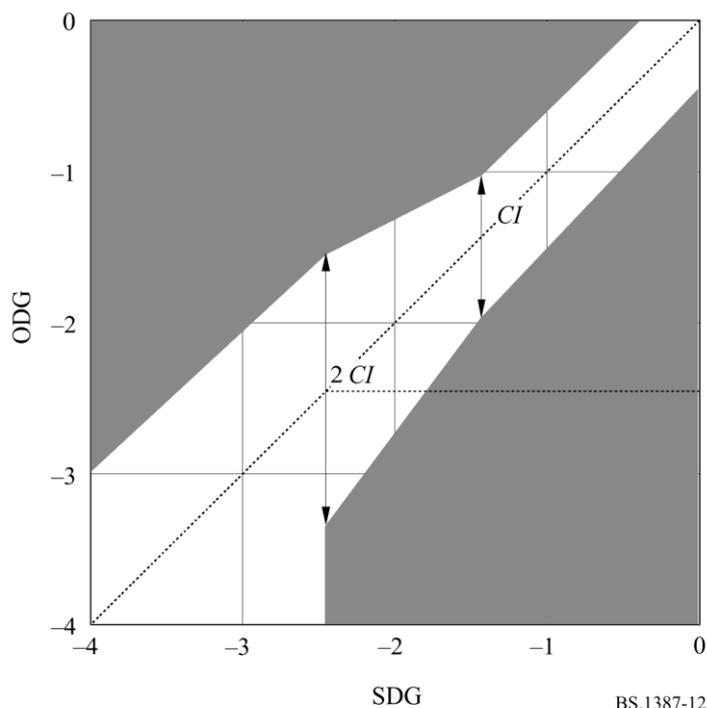
Критерий AES, введенный для связи точности модели с точностью испытания с прослушиванием, определяется в соответствии со следующим выражением:

$$AES = 2 * \sqrt{\frac{\sum ((ODG - SDG) / CI)^2}{N}}, \quad \text{если } CI < 0,25, \text{ то } CI = 0,25,$$

где CI – доверительный интервал.

Схема допуска разработана таким образом, чтобы на верхнем и нижнем концах шкалы ухудшения качества звука допускались различные отклонения ODG от SDG. Допустимый диапазон связан с доверительными интервалами испытаний с прослушиванием. Этот диапазон ограничен минимальным значением 0,25 балла. Для оценки качества метода измерения использовалось расстояние от ODG вне схемы допуска до схемы допуска.

РИСУНОК 12

Схема допуска, доверительный интервал $CI \geq 0,25$ 

4.1 Сравнение значений SDG и ODG

Объективные измерения были разделены на три этапа. На этапе 1 все 84 исследуемых элемента были не известны никому, кроме отборочной комиссии. На этапе 2 была опубликована информация по 52 элементам. Информация содержала как значения SDG, так и фактические аудиофрагменты. На этапе 3 эти сведения использовались для оптимизации рабочих характеристик версий метода. Отметим, что на этапе 3 были протестированы четыре дополнительные версии помимо версий, протестированных на этапе 1. Представленные значения SDG были рассчитаны на основе данных, полученных от 75 квалифицированных участников.

Существует много разных способов оценить, насколько хорошо ODG отражают SDG. К сожалению, ни одно значение не демонстрирует полное соответствие. На самом деле приходится смотреть с нескольких точек зрения. Корреляция представлена в пункте 4.2, а AES – в пункте 4.3. Модель В3 стала одной из моделей, протестированных МСЭ-R в 1996 году, и было решено, что новые версии следует сравнивать с этой версией.

4.2 Корреляция

Коэффициенты корреляции, полученные на этапах 1 и 3, представлены на рисунке 13 (84 элемента) и рисунке 14 (32 элемента).

РИСУНОК 13
Корреляция между SDG и ODG
Включены все 84 элемента

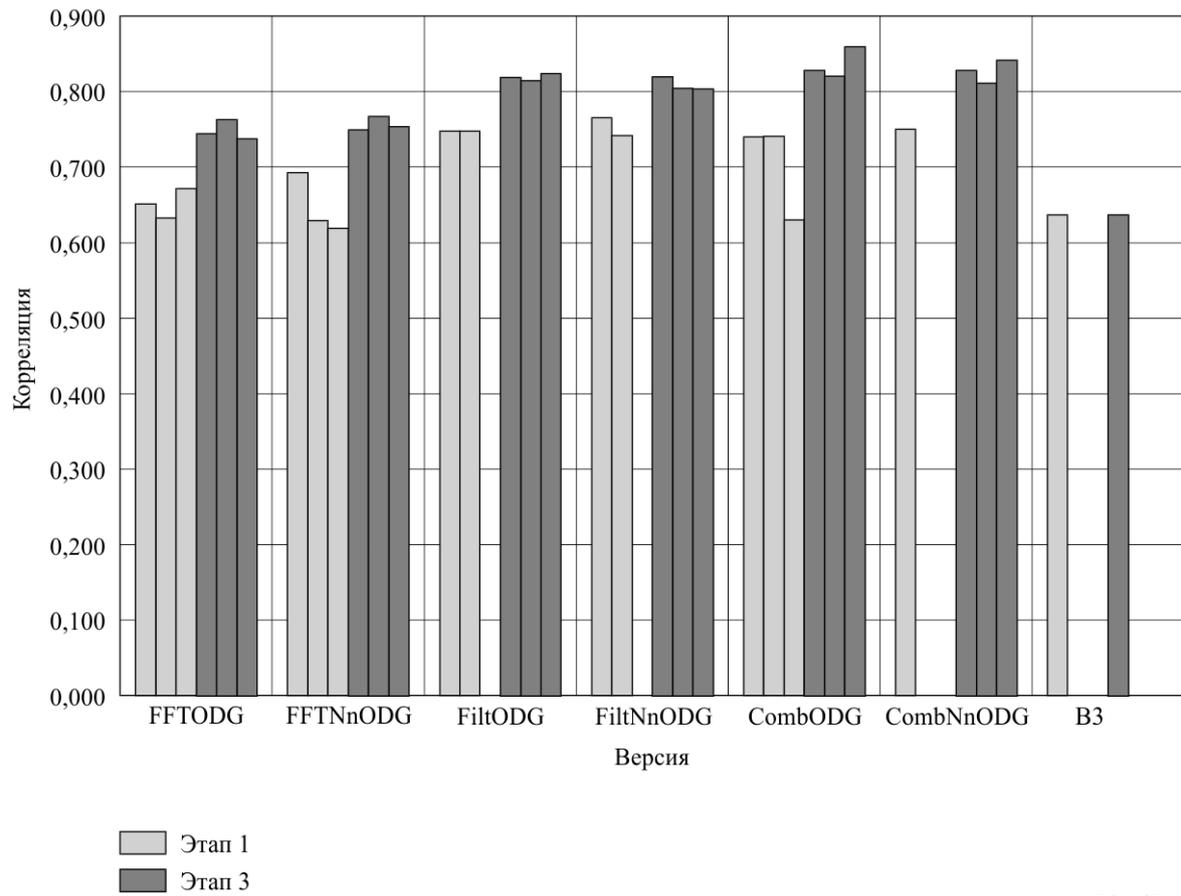
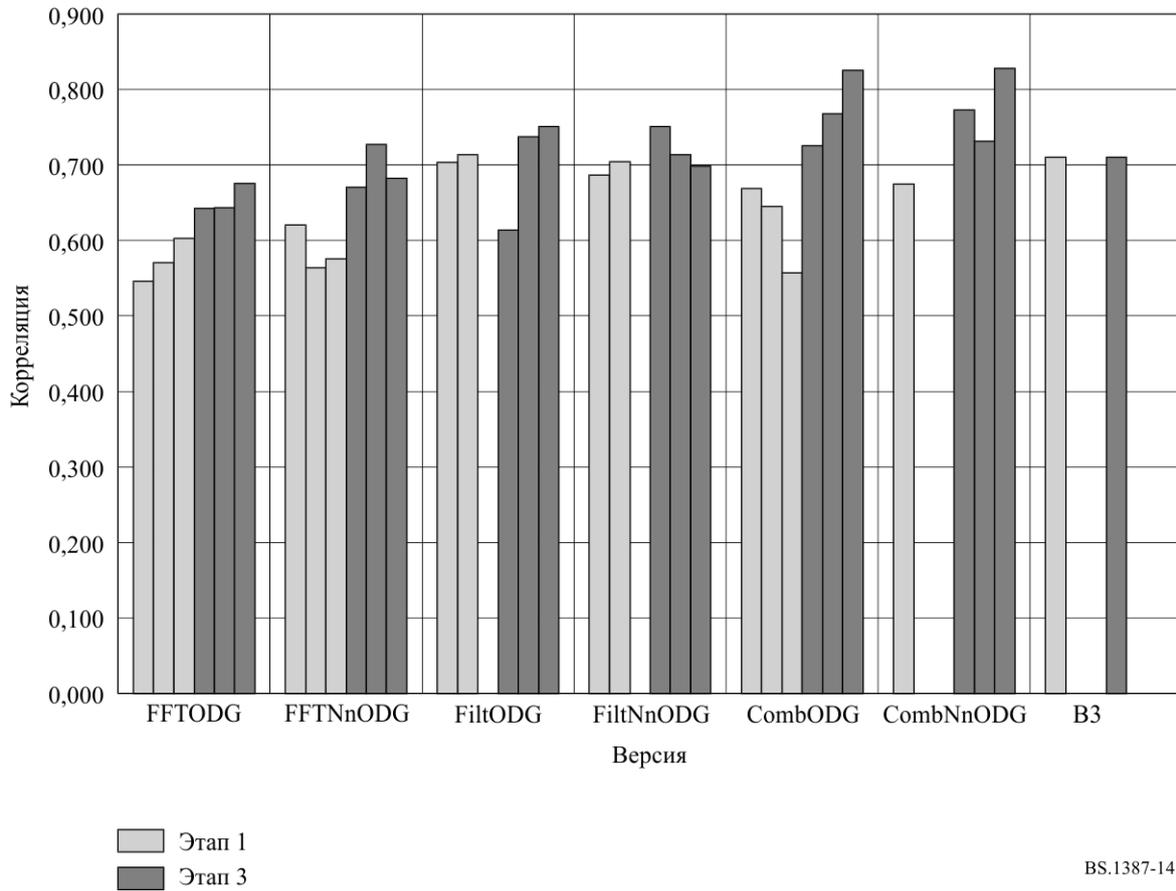


РИСУНОК 14
Корреляция между SDG и ODG
 Включены 32 неопубликованных элемента



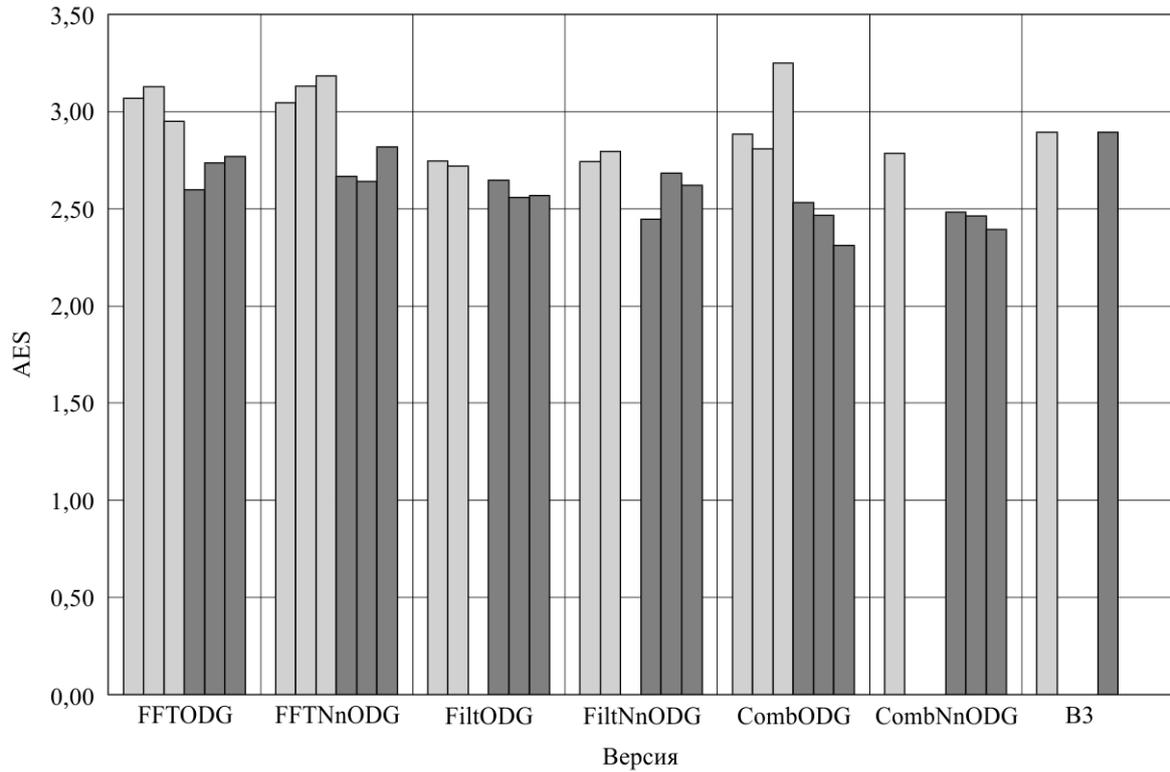
BS.1387-14

4.3 Абсолютная оценка ошибки (AES)

Модель, которая в среднем выдает значения ODG в пределах доверительного интервала SDG, получает значение AES, близкое к 2. Обзор значений AES представлен на рисунках 15–16.

РИСУНОК 15

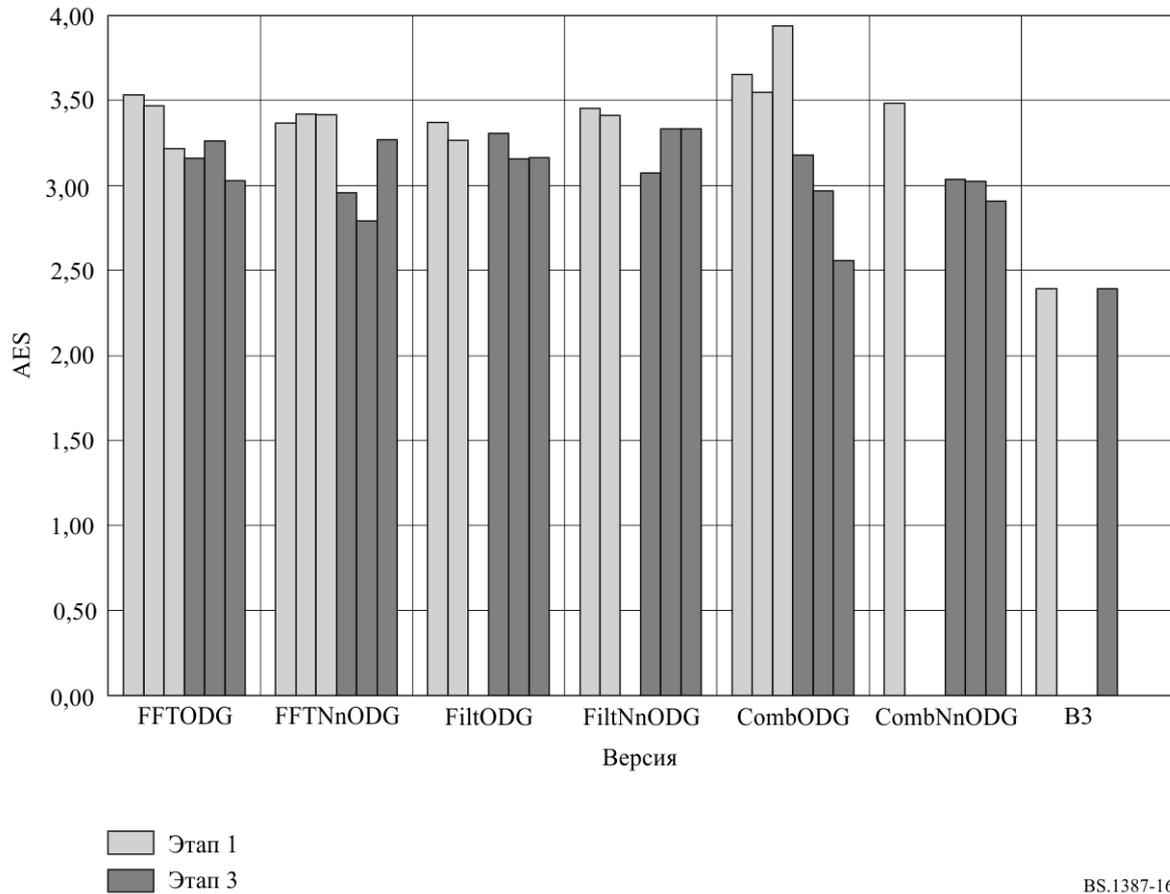
АЭС для разных версий
Включены все 84 элемента



Этап 1
Этап 3

BS.1387-15

РИСУНОК 16
АЭС для разных версий
Включены 32 неопубликованных элемента



BS.1387-16

4.4 Сравнение ODG с доверительным интервалом

Как и следовало ожидать, большинство версий работали одинаково. На собрании было представлено много сюжетов, но в данной Рекомендации их перечень ограничен. Более подробную информацию можно найти в полном отчете о проверочных испытаниях.

На рисунках 17–22 показаны средние значения SDG, доверительный интервал и ODG для 32 неопубликованных элементов для модели B3 и версий моделей FFTNnODG1 и CombNnODG3.

РИСУНОК 17

Модель В3: график средних значений SDG, доверительного интервала и ODG для 32 неопубликованных элементов

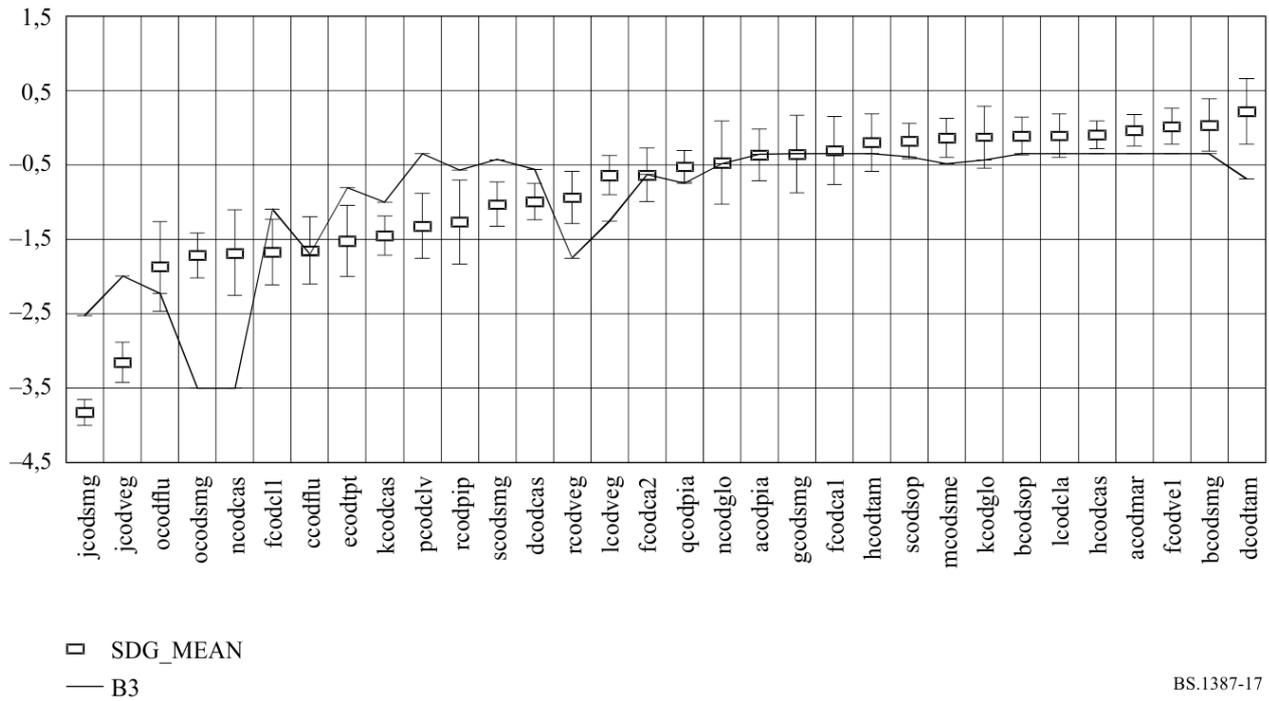


РИСУНОК 18

FFTNnODG1: график средних значений SDG, доверительного интервала и ODG после этапа 3 для 32 неопубликованных элементов

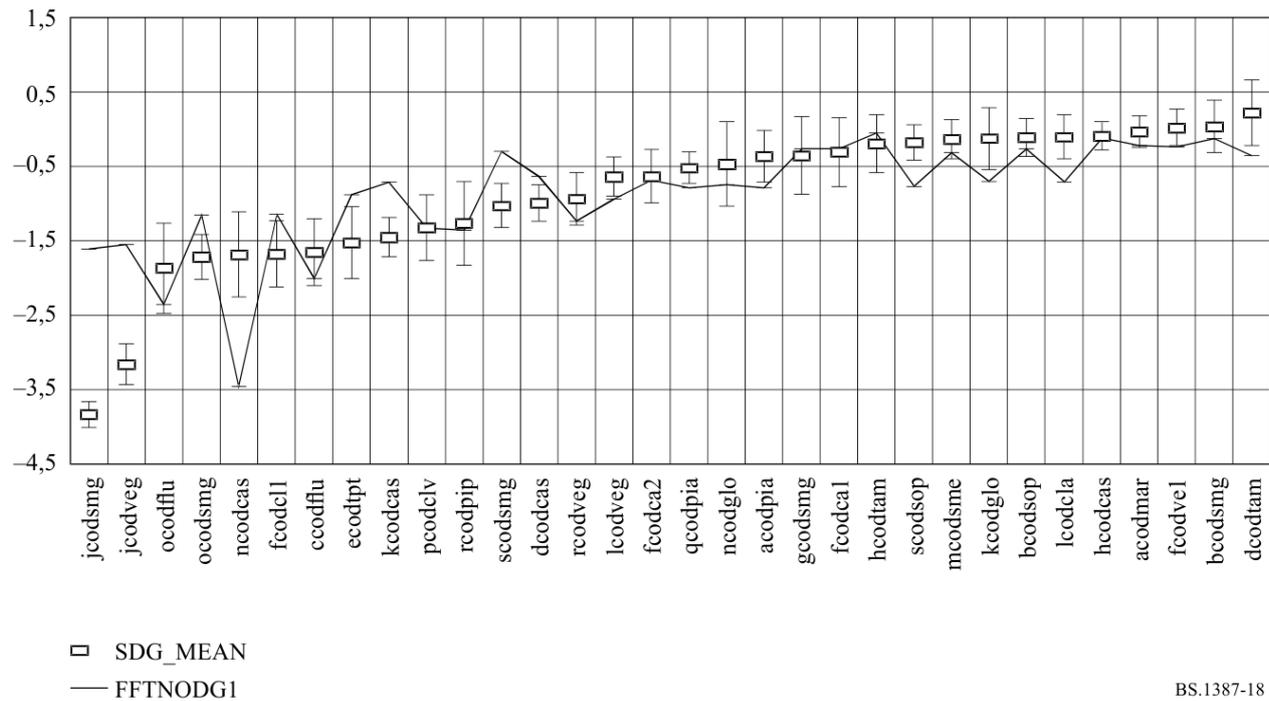
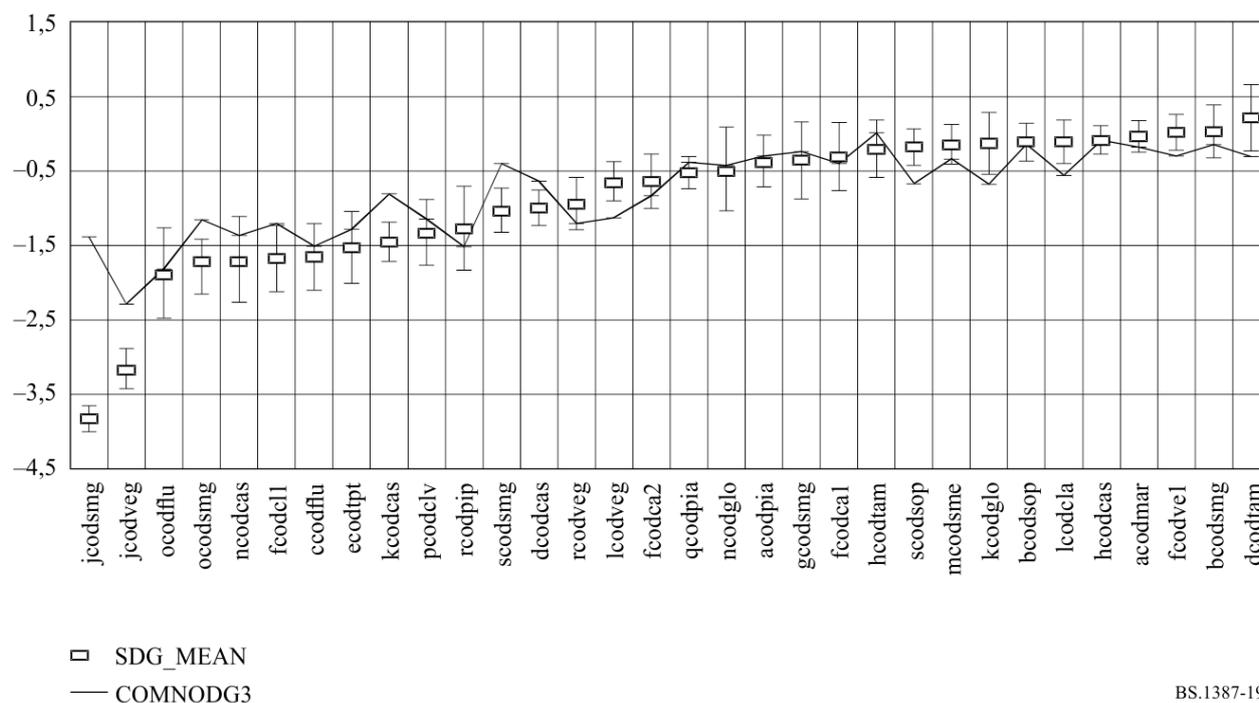


РИСУНОК 19

CombNnODG3: график средних значений SDG, доверительного интервала и ODG после этапа 3 для 32 неопубликованных элементов



Аналогичные графики, но в данном случае для всех 84 элементов на этапе 3, приведены на рисунках 20 и 21. Кроме того, на рисунке 22 иллюстрируются показатели версии CombNnODG3.

РИСУНОК 20

Модель В3: график средних значений SDG, доверительного интервала и ODG для всех 84 элементов

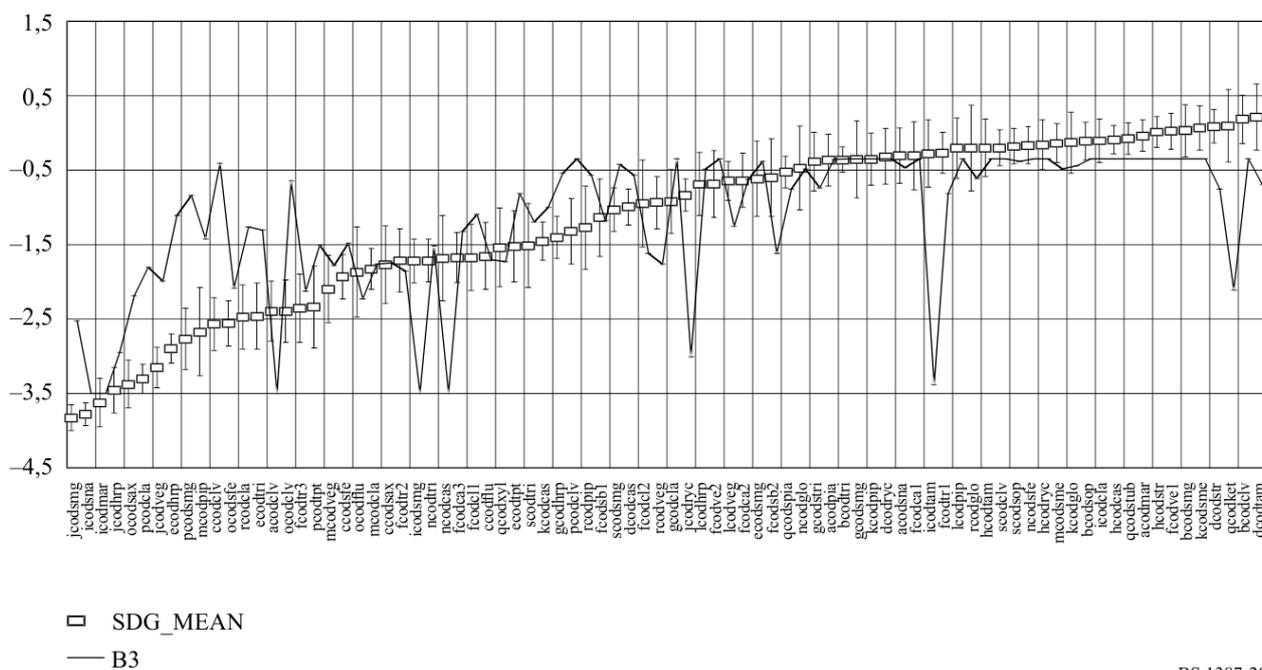
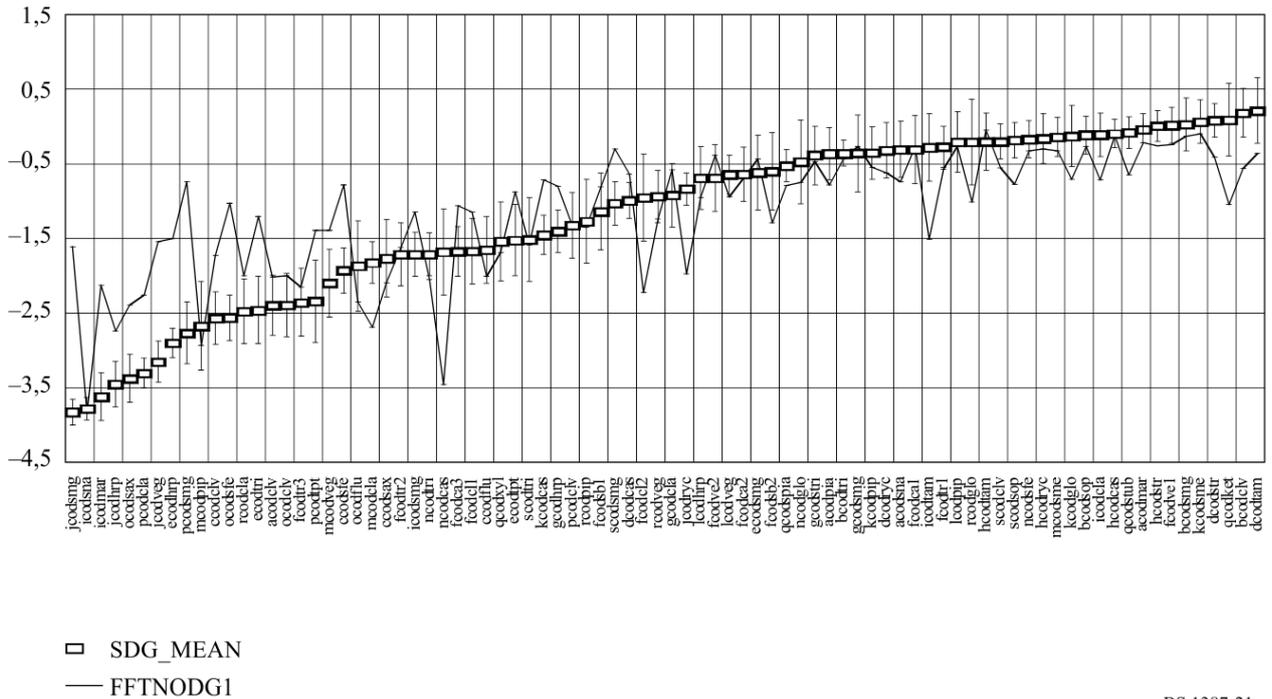


РИСУНОК 21

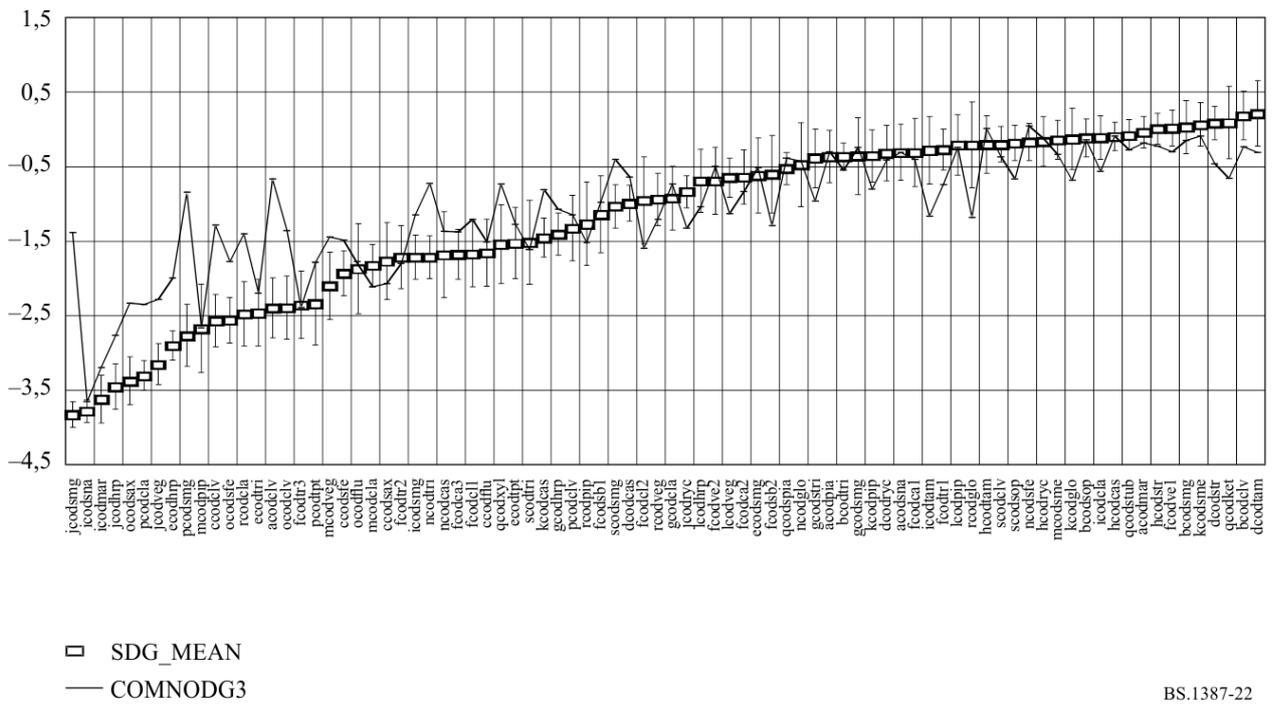
FFTNnODG1: график средних значений SDG, доверительного интервала и ODG на этапе 3 для всех 84 элементов



BS.1387-21

РИСУНОК 22

CombNnODG3: график средних значений SDG, доверительного интервала и ODG на этапе 3 для всех 84 элементов



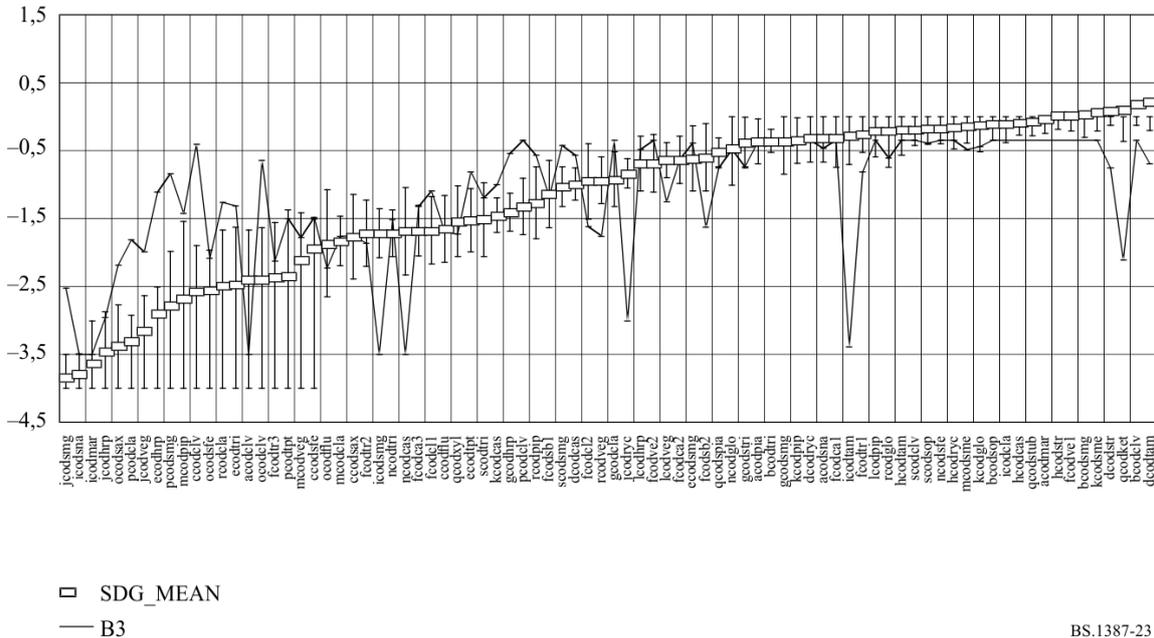
BS.1387-22

4.5 Сравнение ODG с интервалом допуска

МСЭ-R определил целевое требование к пользователям, которое может быть преобразовано в интервал допуска. В отношении звука более высокого качества предъявляются более строгие целевые требования, а в отношении звука более низкого качества – более мягкие. На следующих рисунках показаны характеристики по этому параметру модели В3 и версий модели FFTNnODG1 и CombNnODG3 для всех 84 элементов на этапе 3.

РИСУНОК 23

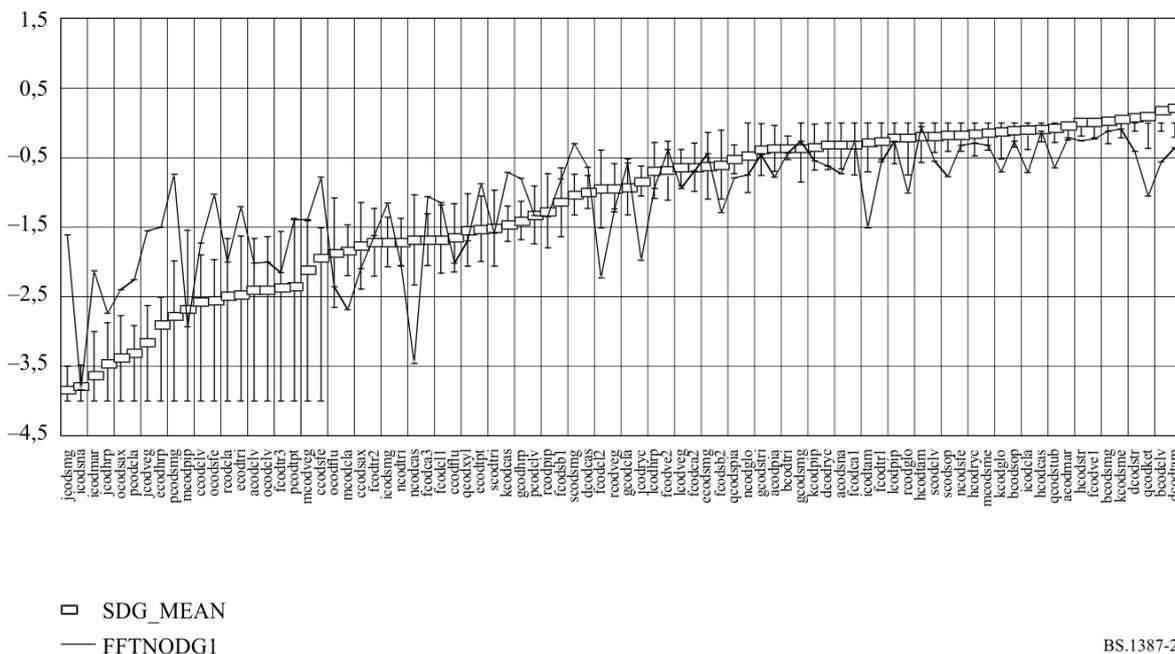
В3: график среднего значения SDG, интервала допуска и ODG на этапе 3 для всех 84 элементов



BS.1387-23

РИСУНОК 24

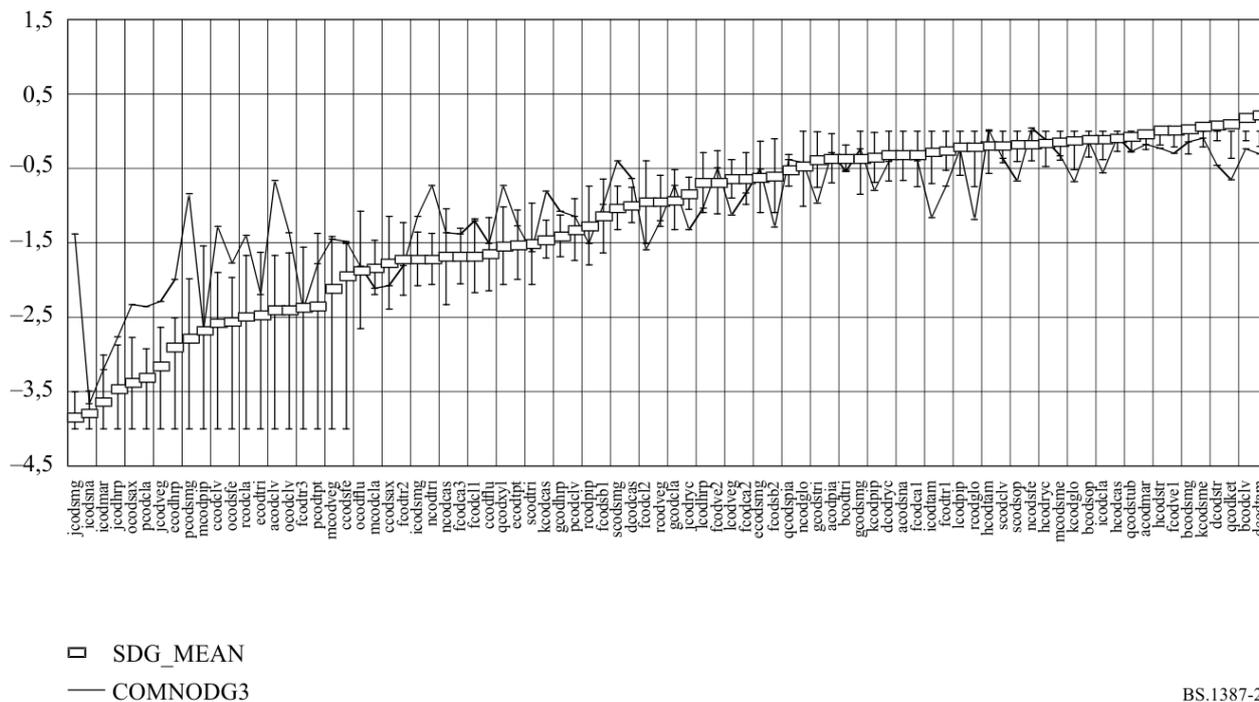
FFTNnODG1: график среднего значения SDG, интервала допуска и ODG на этапе 3 для всех 84 элементов



BS.1387-24

РИСУНОК 25

CombNnODG3: график среднего значения SDG, интервала допуска и ODG на этапе 3 для всех 84 элементов



5 Выбор оптимальных версий модели

На площадке для проведения объективных испытаний было представлено 18 версий модели: шесть версий на основе БПФ, шесть версий с использованием адаптированного к восприятию набора фильтров и шесть версий с использованием комбинации БПФ и набора фильтров. Группа БПФ предназначена для схемы измерения, которую можно реализовать в режиме реального времени; для остальных двух групп требуются более сложные вычисления, и ожидается, что они дадут более высокую точность. Шесть версий модели в каждой группе можно разделить на две подгруппы: в одной используются нейронные сети, а в другой нет. Для сравнения также включены характеристики эталонной модели от апреля 1996 года (V3) на основе базы данных 3.

5.1 Критерии предварительного отбора, основанные на корреляции

- Характеристики 18 версий модели и эталонной модели V3 оценивались с применением двух наборов данных, которые в основном не использовались для обучения моделей (этап 1, вторая часть этапа 3 и CRC для обучения не использовались).
- База данных 3 была создана специально для оценки моделей восприятия. Примерно половина этой базы данных использовалась для обучения моделей. Корреляция между субъективными и объективными результатами по остальным пунктам (DB3_2nd) использовалась для оценки объективных моделей.
- База данных CRC была создана CRC для оценки схем перцепционного кодирования звука. Эта база данных не использовалась для обучения моделей восприятия. Для оценки объективных моделей использовалась корреляция между субъективными и объективными результатами по всем элементам.

Результаты базы данных 3 приведены в таблицах 24 и 25 (корреляция и AES).

На этапе предварительного отбора были учтены все входные данные, определяющие характеристики всех версий модели, с высоким весом в DB3. На основании общего сравнения было решено провести дальнейшее сравнение двух версий модели в каждой группе, которая оказалась лучшей. В таблице 24 показано соотношение этих шести версий модели (трех пар).

ТАБЛИЦА 24

Корреляция между SDG и ODG

	FFTNnODG1	FFTNnODG2	FiltODG2	FiltODG3	CombNnODG3	CombODG3	B3
DB3_2nd	0,671	0,728	0,738	0,751	0,828	0,826	0,710
CRC	0,837	0,779	0,862	0,839	0,851	0,777	0,656

ТАБЛИЦА 25

Абсолютная оценка ошибки

	FFTNnODG1	FFTNnODG2	FFTOG2	FFTOG3	CombNnODG3	CombODG3	B3
DB3_2nd	2,96	2,79	3,16	3,16	2,91	2,56	2,39
CRC	1,55	1,85	1,61	1,67	1,61	1,90	2,78

Две версии модели на основе БПФ демонстрируют разные характеристики по двум базам данных. Для решения о выборе одной из них требуются дополнительные критерии. Принимая во внимание, что база данных CRC была абсолютно неизвестна, похоже, что FFTNnODG1 работает немного лучше.

Комбинированные версии модели дают лучшие результаты на обеих базах данных по сравнению с версиями на основе БПФ. Однако они показывают худшие характеристики для базы данных CRC по сравнению с версиями на основе набора фильтров. Версии на основе набора фильтров можно рассматривать как частный случай комбинированных версий модели, когда весовые коэффициенты выходных переменных версии на основе БПФ равны нулю. Поэтому комбинированные версии моделей предпочтительны.

5.2 Анализ количества выбросов

Точность субъективных данных, полученных в результате прослушивания, выражается через 95%-ный интервал вокруг среднего показателя по нескольким слушателям. Характеристики версий модели оценивались также путем рассмотрения выбросов. Элемент считается выбросом, если разница между субъективными и объективными данными по нему превышает доверительный интервал более чем в два раза.

В таблице 26 показано количество выбросов для шести версий модели DB3. "Повышенная чувствительность" означает, что версия модели указывает на более низкое качество звука, чем субъективная оценка, а "пониженная" – наоборот.

ТАБЛИЦА 26

Количество выбросов

	FFTNnODG1	FFTNnODG2	FiltODG2	FiltODG3	CombNnODG3	CombODG3
Повышенная чувствительность	10	4	4	4	3	5
Пониженная чувствительность	13	13	11	13	12	14
Всего	23	17	15	17	15	19

Судя по количеству выбросов, версия модели FFTNnODG2 работает лучше, чем версия модели FFTNnODG1. Версия модели CombNnODG3 показывает лучшие характеристики среди более точных версий.

5.3 Анализ величины выбросов

ТАБЛИЦА 27

Элементы с расхождением между прогнозом и SDG более чем на 1,0 балла

Количество выбросов	14	4	4	12	12	9	6
Версия модели	B3	CombODG3	CombNnODG3	FFTNnODG1	FFTNnODG2	FiltODG2	<u>FiltODG3</u>
Элемент	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg
	qcodket	qcodket	pcodsmg	qcodket	qcodket	rcodpip	rcodpip
	pcodsmg	pcodsmg	ccodclv	pcodsmg	pcodsmg	qcodket	qcodket
	pcodcla	icodmar	acodclv	ocodsfe	pcodcla	pcodsmg	pcodsmg
	ocodclv			ncodcas	ocodsfe	ocodsax	ocodsax
	ncodcas			mcodcla	ncodglo	kcodpip	fcodsb2
	mcodpip			jcodveg	kcodcas	jcodveg	
	jcodyrc			jcodyrc	jcodveg	fcodsb2	
	icodtam			icodtam	jcodyrc	fcodcl2	
	icodsmg			icodmar	icodtam		
	fcodsb2			fcodcl2	ecodhrp		
	ecodhrp			ecodhrp	ccodclv		
	ccodclv						
	acodclv						

ТАБЛИЦА 28

Элементы с расхождением между прогнозом и SDG более чем на 1,5 балла

Количество выбросов	8	2	2	3	2	2	2
Версия модели	B3	CombODG3	CombNnODG3	FFTNnODG1	FFTNnODG2	FiltODG2	<u>FiltODG3</u>
Элемент		jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg	jcodsmg
	qcodket		pcodsmg				
	pcodsmg	pcodsmg		pcodsmg	pcodsmg		
						pcodsmg	pcodsmg
				ncodcas			
	ncodcas						
	jcodyrc						
	icodtam						
	icodsmg						
	ccodclv						
	acodclv						

6 Выводы

В соответствии с решениями, принятыми в ходе разработки, из предложенных 18 версий модели были выбраны две, отвечающие установленным требованиям к применению объективных методов измерения. Была определена версия, имеющая низкую сложность, предназначенная для обеспечения экономичной реализации в режиме реального времени, и версия с более высокой точностью, не обязательно работающая в режиме реального времени. В процессе отбора применялись и анализировались вышеперечисленные критерии.

В качестве версии реального времени была выбрана модель на основе БПФ, называемая FFTNnODG1, по следующим причинам.

Ни одна из версий на основе БПФ не показала существенного преимущества перед другими ни по одному из перечисленных выше критериев. Что касается коэффициентов корреляции между ODG и SDG, то каждая из проверенных версий имели свои преимущества и недостатки с точки зрения количества и величины выбросов. Тем не менее следует отметить, что версия FFTNnODG1 достигла наилучшей корреляции по базе данных CRC'97, которая была абсолютно неизвестна (0,837).

В качестве версии с более высокой точностью предпочтение было отдано комбинированной версии, включающей БПФ и набор фильтров, поскольку такой подход также охватывает подмножество моделей, основанных только на наборе фильтров и, следовательно, должен потенциально показывать лучшие результаты. Всего на выбор предлагалось шесть разных версий комбинированной модели. Выбранная версия CombNnODG3 показала меньше выбросов, а также более высокую корреляцию ($r = 0,851$ для CRC'97), чем другие версии. Корреляция этой версии для полной базы данных 3 имела тот же порядок величины, что и версии CombODG3, но она показала лучшую корреляцию по сравнению с другими версиями.

Прилагаемый документ 2 к Приложению 2

Описание справочных баз данных

1 Введение

При разработке методики объективного измерения воспринимаемого качества звука для обучения и проверки использовался ряд баз данных.

Некоторые из перечисленных баз данных содержат данные для прослушивания как через наушники, так и через громкоговорители, а другие – только данные для прослушивания через наушники. В случае баз данных, содержащих отдельные наборы данных для воспроизведения через громкоговорители и через наушники, использовались только данные для наушников.

Элемент определяется как аудиофрагмент, используемый при субъективной оценке. Под условием понималось условие единичного дефекта качества. При всех условиях экспериментов, за исключением исследований с применением DB2 и DB3, использовались все элементы. DB3 использовалась частично для обучения и частично для проверки (52 из 84 элементов использовались для обучения на втором этапе проверки).

Обучение

– MPEG90

- Среднее значение SDG по каждому элементу довольно равномерно охватывало диапазон от 0,0 до -4,0.
- Отчет об испытаниях ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio, документ MPEG90/N0030, октябрь 1990 года.

- MPEG91
 - По крайней мере 88% средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $0,1$ до $-3,8$.
 - Отчет об испытаниях ISO/IEC JTC 1/SC 2/WG 11 MPEG/Audio, документ MPEG91/N0010, июнь 1991 года.
- ITU92DI
 - 80% средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $0,1$ до $-3,4$.
- ITU92CO
 - По крайней мере 96% средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $0,2$ до $-2,4$.
- ITU93
 - Большинство средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $-0,1$ до $-2,3$. Между данными двух лабораторий не было существенных расхождений.
 - Grusec *et al* [1997].
- MPEG95
 - По крайней мере 63% средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $-0,2$ до $-3,8$.
 - Meares and Kim [1995].
- EIA95
 - По крайней мере 93% средних значений SDG по каждому элементу были выше $-2,0$, а их диапазон составил от $0,1$ до $-3,7$.
 - Grusec *et al* [1997].
- DB2
 - Не все элементы использовались при всех условиях.

Проверка

- DB3
 - Не все элементы использовались при всех условиях.
- CRC97
 - Среднее значение SDG по каждому элементу довольно равномерно охватывало диапазон от $0,1$ до $-3,6$.
 - Soulodre *et al* [1998].

В следующих разделах описаны элементы, включенные в различные базы данных, и применяемые условия.

2 Элементы, содержащиеся в базах данных

Элемент	MPEG90	MPEG91	ITU92DI	ITU92CO	ITU93	MPEG95	EIA95	DB2	DB3	CRC97
Аккордеон/треугольник		*								
Оса Джинджер			*	*	*					
Волынка						*			*	
Волынка 2								*		
Вас-кларнет								*		
Вас-гитара	*		*	*						
Басовый синтезатор	*									
Кармен		*								
Кастаньеты	*		*	*	*	*		*	*	
Кларнет					*		*	*	*	*
Кларнет моно								*		
Кларнет 2								*		
Клаве									*	
Dalarnas Spelmansförbund "Trettondagsmarschen"			*	*						
Dire Straits "Ride Across the River"			*				*			*
Контрабас										*
Барабан								*		
Пиротехнические средства	*									
Флейта									*	
Джордж Дюк		*								
Колокольчики	*	*				*	*	*	*	
Клавесин			*	*	*	*	*	*	*	*
Рожок								*		
Литавры									*	
Маримба								*	*	
Мужская речь, англ., моно								*		
Музыка и шум дождя							*			*
Труба с надетой сурдинкой							*			
Орнетт Коулман	*	*	*	*						
Real Jap							*			
Ударные		*						*		
Фортепьяно, Шуберт									*	
Камертон-дудка						*			*	*
Равель, "Ферия"			*							
Рай Кудер								*	*	
Рай Кудер, моно										
Саксофон									*	
Малый барабан									*	
Сопрано, Моцарт									*	
Женская речь, англ.						*			*	
Женская речь, нем.									*	
Мужская речь, англ.	*	*							*	
Мужская речь, нем.			*	*	*				*	
Стравинский, октет для духовых инструментов			*	*						
Струнные								*	*	
Струнные, моно								*		
Сюзанна Вега "Tom's Diner"	*	*	*	*				*	*	*
Сюзанна Вега со звуком бьющегося стекла							*			
Тамбурин		*						*	*	
Трейси Чепмен	*									
Треугольник			*	*				*	*	
Труба								*	*	*
Труба, Гайдн	*									
Туба								*	*	
De sålde sina hemman (соло скрипки)					*					
Журчание воды							*			
Ансамбль духовых инструментов								*		
Ксилофон									*	

3 Условия экспериментов

Для всех значений битовой скорости с указанием "кбит/с стерео" указана общая битовая скорость; например, "256 кбит/с стерео" означает, что в общей сложности для обоих каналов стереосигнала выделена полоса 256 кбит/с. Если не указано иное, то "стерео" означает независимое кодирование каналов.

3.1 MPEG90

Три значения битовой скорости: 64 кбит/с моно, 192 кбит/с и 256 кбит/с стерео; для этой базы данных был доступен не весь материал.

- Musicam
- SB-ADPCM

3.2 MPEG91

Три значения битовой скорости: 64 кбит/с моно, 192 кбит/с и 256 кбит/с стерео.

- MPEG1 уровень I
- MPEG1 уровень II
- MPEG1 уровень III
- MUSICAM
- ASPEC
- NICAM

3.3 ITU92DI

Пять кодеков для распределения: 240 кбит/с стерео.

Каждый элемент обрабатывался одним и тем же кодеком три раза подряд с понижением уровня на 0,1 дБ перед каждым проходом.

- MPEG1 уровень II
- MPEG1 уровень III
- Dolby AC-2
- Aware
- NHK

3.4 ITU92CO

Шесть кодеков для доставки: 360 кбит/с стерео. Каждый элемент обрабатывался одним и тем же кодеком три раза подряд с понижением уровня на 0,1 дБ перед каждым проходом.

- MPEG1 уровень II
- MPEG1 уровень III
- Dolby AC-2
- Dolby Low-Delay
- Aware

3.5 ITU93

Многокаскадные конфигурации кодеков MPEG1 уровня II:

- только кодек радиопередачи со скоростью 256 кбит/с стерео;
- только кодек радиопередачи со скоростью 192 кбит/с стерео (совместное стереокодирование);
- восемь кодеков доставки со скоростью 360 кбит/с, за которыми следует один кодек радиопередачи со скоростью 256 кбит/с, все стерео;
- восемь кодеков доставки со скоростью 360 кбит/с, за которыми следует один кодек радиопередачи со скоростью 192 кбит/с, все стерео;
- пять кодеков доставки со скоростью 360 кбит/с, за которыми следуют три кодека распределения со скоростью 240 кбит/с и один кодек радиопередачи со скоростью 256 кбит/с, все стерео;

- пять кодеков доставки со скоростью 360 кбит/с, за которыми следуют три кодека распределения со скоростью 240 кбит/с и один кодек радиопередачи со скоростью 192 кбит/с, все стерео.

3.6 MPEG95

Реализации кодеков (64 кбит/с):

- были выбраны 22 варианта из более широкого набора доступных методов кодирования, состоящего из шести кодеков, реализующих подмножество из четырех частотно-временных моделей с низким разрешением и 17 частотно-временных моделей с высоким разрешением;
- в испытаниях приняли участие компании AT&T, Институт Фраунгофера, Sony, GCL, RAI/Alcatel и Philips;
- все элементы представляли собой монофонические записи, представленные в бинауральном формате.

3.7 EIA95

- Eureka 147/MPEG1 уровень II #1 224 кбит/с стерео (совместное стереокодирование)
- Eureka 147/MPEG1 уровень II #2 192 кбит/с стерео (совместное стереокодирование)
- AT&T/Lucent 160 кбит/с стерео
- AT&T/Lucent/Amati #1 128 кбит/с стерео
- AT&T/Lucent/Amati #2 160 кбит/с стерео
- VOA/JPL 160 кбит/с стерео
- USADR-FM #1 128–256 кбит/с стерео (переменная битовая скорость)
- USADR-FM #2 128–256 кбит/с стерео (переменная битовая скорость)
- USADR-AM9 96 кбит/с стерео

3.8 DB2

- MPEG1 уровень II 256 кбит/с стерео, этапы 1, 3, 5, 7 и 9
- Dolby AC2 256 кбит/с стерео, этапы 1, 3, 5, 7 и 9
- MPEG1 уровень II 192 кбит/с стерео (совместное стереокодирование)
- MPEG1 уровень II 64 кбит/с моно
- MPEG2 уровень II 64 кбит/с моно
- MPEG1 уровень II 384 кбит/с стерео
- MPEG1 уровень III 128, 160, 192 кбит/с, все стерео
- APT-X 256 и 384 кбит/с, оба стерео
- Искажение квантования
- Аналоговая запись, этапы 1, 2, 3
- Обрезка

3.9 DB3

- NICAM
- MiniDisc и MiniDisc + уровень II 192 кбит/с стерео (совместное стереокодирование)
- Dolby AC2 256 кбит/с стерео, этапы 1, 3, 5, 7 и 9
- MPEG1 уровень II выборка из базы данных Swisscom, > 192 кбит/с стерео
- MPEG1 уровень III 128 и 160 кбит/с, оба стерео (совместное стереокодирование)

–	MPEG AAC	128 кбит/с стерео (совместное стереокодирование)
–	MPEG уровень III	128 + уровень II, 384 + уровень II, 224 кбит/с, все стерео
–	Dolby AC3	256 кбит/с стерео
–	Dolby AC3	256 + MPEG уровень II, 224 кбит/с, оба стерео
–	Искажение квантования	
–	THD	
–	Шум	

3.10 CRC97

–	AT&T PAC	64, 96, 128 и 160 кбит/с, все стерео
–	Dolby AC3	128, 160 и 192 кбит/с, все стерео
–	MPEG1 уровень II программн.	128, 160 и 192 кбит/с, все стерео
–	MPEG1 уровень II аппаратн. (ITIS)	96, 128, 160, 192 кбит/с, все стерео
–	MPEG4 AAC	96 и 128 кбит/с, оба стерео
–	MPEG1 уровень III	128 кбит/с стерео

4 Элементы из DB2 и DB3 для каждого условия

4.1 DB2

Номер условия		Элементы
Опытная площадка I, NHK, Япония		
Уровень II, 256 кбит/с 1 этап	CO13	CLA,RYC,SB1,STR
3 этапа	CO11	CLA,RYC,SB1,STR
5 этапов	CO19	CLA,RYC,SB1,STR
7 этапов	CO18	CLA,RYC,SB1,STR
9 этапов	CO15	CLA,RYC,SB1,STR
NBC (Dolby AC2) 1 этап	CO1A	CAS,RYC,STR,WIN
3 этапа	CO12	CAS,RYC,STR,WIN
5 этапов	CO17	CAS,RYC,STR,WIN
7 этапов	CO16	CAS,RYC,STR,WIN
9 этапов	CO14	CAS,RYC,STR,WIN

Номер условия		Элементы
Опытная площадка II, DR, Дания		
Уровень II, 256 кбит/с 1 этап	CO2B	CLA,RYC,SB1,STR
Уровень II, 192 кбит/с, совместное стереокодирование	CO25	CLA,RYC,SB1,STR
Уровень II, 64 кбит/с моно	CO27	MLA,MPE,MTR,MYC
NBC (Dolby AC2) 5 этапов	CO29	CAS,RYC,STR,WIN
MPEG2/L2 LSF	CO22	MLA,MPE,MTR,MYC
Аналоговый 1	CO23	PER
Аналоговый 2	CO2A	PER
Аналоговый 3	CO28	PER
Ошибки 1	CO24	GLO,HRN,TRI
Ошибки 2	CO21	GLO,HRN,TRI
Обрезка	CO26	BAS,CL2,TUB

Опытная площадка III, NRK, Норвегия

Уровень II, 384 кбит/с	CO34	CLA,RYC,SB1,STR
Уровень II, 256 кбит/с 1 этап	CO31	CLA,RYC,SB1,STR
NBC (Dolby AC2) 5 этапов	CO3B	CAS,RYC,STR,WIN
Уровень III (ASPEC3), 192 кбит/с	CO32	CLA,STR,TAM,VEG
Уровень III (ASPEC3), 128 кбит/с	CO39	CLA,STR,TAM,VEG
Уровень III (ASPEC3), 160 кбит/с	CO3A	CLA,STR,TAM,VEG
APT-X, 256 кбит/с	CO33	HAR,SB2,STR,TPT
APT-X, 384 кбит/с	CO36	HAR,SB2,STR,TPT
Искажение квантования 1	CO35	DRU
Искажение квантования 2	CO37	DRU
Искажение квантования 3	CO38	DRU

Исследуемые элементы

STR	Шведская народная музыка, запись с шумоподавлением (SR), использовалась ранее
SB1	Волынки, запись SR
SB2	Волынки, запись SR
CLA	Кларнет, SQUAM 16/2
TAM	Бубен, запись SR, использовалась ранее
WIN	Стравинский, ансамбль духовых инструментов, использовалась ранее
TPT	Труба, SQUAM 21/2
HAR	Ансамбль "Арлекин", запись BBC G 49/17
VEG	Сюзанна Вега, необновленный исходный файл, использовалась ранее
CAS	Кастаньеты, SQUAM 27
SPE	Немецкая речь, SQUAM 54
RYC	Рай Кудер, компакт-диск: ДЖАЗ трек 11 (0,25-0,47)
PER	Ударные, японская бас-маримба, компакт-диск: Sony/CBS 32DC 5027
HRN	Рожок, SQUAM 23/2
GLO	Колокольчики, SQUAM 35/1, использовалась ранее
TRI	Треугольник, SQUAM 32/2
DRU	Барабаны, SQUAM 28
CL2	Кларнет, SQUAM 16/2
BAS	Бас-кларнет, SQUAM 17
TUB	Туба, SQUAM 24
MPE	Моно-микс SPE
MTR	Моно-микс STR
MLA	Моно-микс CLA
MYC	Моно-микс RYC

Исследуемый элемент	Кларнет	Кларнет моно	Рай Кудер	Рай Кудер моно	Волынки 1	Струнные	Струнные моно	Кастаньеты	Ансамбль духовых инструментов	MPE моно	Тамбурин	Сюзанна Вега	Клавесин	Волынки 2	Труба	Барабан	Ударные	Колокольчики	Рожок	Треугольник	Бас-кларнет	Кларнет 2	Туба
MPEG1 уровень 2, 256 кбит/с, 1 этап	X		X		X	X																	
MPEG1 уровень 2, 256 кбит/с, 3 этапа	X		X		X	X																	
MPEG1 уровень 2, 256 кбит/с, 5 этапов	X		X		X	X																	
MPEG1 уровень 2, 256 кбит/с, 7 этапов	X		X		X	X																	
MPEG1 уровень 2, 256 кбит/с, 9 этапов	X		X		X	X																	
Dolby AC2, 256 кбит/с, 1 этап			X			X		X	X														
Dolby AC2, 256 кбит/с, 3 этапа			X			X		X	X														
Dolby AC2, 256 кбит/с, 5 этапов			X			X		X	X														
Dolby AC2, 256 кбит/с, 7 этапов			X			X		X	X														
Dolby AC2, 256 кбит/с, 9 этапов			X			X		X	X														
MPEG1 уровень 2, 192 кбит/с, совместное стереокодирование	X		X		X	X																	
MPEG1 уровень 2, 64 кбит/с моно		X		X			X			X													
MPEG2 уровень 2, 64 кбит/с моно		X		X			X			X													
MPEG1 уровень 2, 384 кбит/с	X		X		X	X																	
MPEG1 уровень 3 (ASPEC 3), 192 кбит/с	X					X					X	X											
MPEG1 уровень 3 (ASPEC 3), 128 кбит/с	X					X					X	X											
MPEG1 уровень 3 (ASPEC 3), 160 кбит/с	X					X					X	X											
АРТ-X, 256 кбит/с						X						X	X	X									
АРТ-X, 384 кбит/с						X						X	X	X									
Шум квантования 1																X							
Шум квантования 2																X							
Шум квантования 3																X							
Аналоговая запись, 1 этап																	X						
Аналоговая запись, 2 этапа																	X						
Аналоговая запись, 3 этапа																	X						
Битовые ошибки 1																		X	X	X			
Битовые ошибки 2																		X	X	X			
Обрезка																					X	X	X

4.2 DB3

Исследуемый элемент	Имя	1	2MD	2MDL2	3 1 этап	3 3 этапа	3 5 этапов	3 7 этапов	3 9 этапов	4	5	6 Низкие	6 Высокие	7	8	9	10	11	12	13	
13(1) Флейта	flu							X	X												
16(2) Кларнет	cla			X		X				X	X			X	X						
20(1) Саксофон	sax						X	X													
21(2) Труба	tpt				X	X															
24(2) Туба	tub																		X		
26(1) Клаве	clv	X	X			X	X	X													X
27 Кастаньеты	cas									X		X			X	X			X		
28 Малые барабаны	sna	X																			X
30 Литавры	ket																			X	
32(1/2) Треугольник	tri	X	X	X	X					X											X
35(1/2) Колокольчики	glo										X	X									X
36(1) Ксилофон	xyl																				X
40(1) Клавецин	hrp			X	X						X				X						
49 Женская речь, англ.	sfe						X	X													X
54 Мужская речь, нем.	smg	X	X	X	X	X				X											
60 Фортепиано, Шуберт	pia																		X		X
61 Сопрано, Моцарт	sop	X	X																		
53 Женская речь, нем.	sfg																				
50 Мужская речь, англ.	sme											X			X						
Ref_tam	tam	X														X	X				
Ref_str	str															X	X				
Ref_har	har																				
Рай Кудер	guc									X						X	X				
Сюзанна Вега	veg								X	X	X			X	X						
Камертон-дудка	pip									X	X		X	X							
Маримба	mar	X																			X
Волынка	sb1									X											
Имя		i	b	s	g	e	p	c	o	f	j	r	k	l	m	h	d	q	n	a	

- 1) NICAM
- 2) MD и MD + L2 (192 кбит/с)
- 3) AAC2 (256 кбит/с), 1, 3, 5, 7, 9
- 4) L2 (ST d-b, 192 кбит/с)
- 5) Уровень 2 (256 кбит/с), 8 этапов
- 6) Уровень 3, 128 и 160 кбит/с
- 7) AAC, 128 кбит/с
- 8) L3 (128) + L2 (384) + L2 (224)
- 9) AC3 (256)
- 10) AC3 (256) + L2 (224)
- 11) Шум квантования
- 12) THD
- 13) Шум

Примеры имен:

Эталонные: irefflu
Исследуемые: icodflu

Глоссарий

Абсолютная оценка ошибки (AES)

AES выводится по формуле, разработанной специально для оценки качества результатов, полученных с помощью объективного метода перцептивного измерения. Она учитывает доверительные интервалы средних значений субъективных испытаний с прослушиванием.

Базовое качество звука

Базовое качество звука определяется как глобальный субъективный атрибут, включающий в себя все обнаруженные различия между эталонным сигналом и его обработанной версией.

Запас по кодированию

Запас по кодированию – это параметр качества, измеряющий запас уровня сигнала от неслышимых артефактов кодирования до порогового значения, когда эти артефакты становятся слышимыми.

Выходные переменные модели (MOV)

MOV представляют собой промежуточные выходные значения метода перцептивного измерения. Эти переменные основаны на базовых психоакустических результатах и, следовательно, могут использоваться для дальнейшей характеристики артефактов кодирования.

Объективная оценка разницы (ODG)

ODG является основным выходным параметром метода перцептивного измерения. Она сопоставляется с SDG и представляет собой параметр измерения, определяющий глобальное базовое качество звука. ODG имеет диапазон от 0 до -4.

Офлайн-измерение

Процедура измерения без взаимодействия с текущей передачей программы.

Онлайн-измерение

Процедура измерения, основанная на текущей передаче программы или ее части.

Субъективная оценка разницы (SDG)

При испытании с прослушиванием в соответствии с Рекомендацией МСЭ-R BS.1116 базовое качество звука скрытого эталонного сигнала и его обработанной версии оценивается по пятибалльной шкале ухудшения качества. Оценка разницы определяется как оценка, присвоенная исследуемому сигналу, за вычетом оценки, присвоенной эталонному сигналу. В идеале SDG должна лежать в диапазоне от 0 до -4. Если эталонный сигнал определен неправильно, то SDG будет положительной.

Сокращения

ADB	Average Distorted Block		Средний искаженный блок
AES	Absolute Error Score		Абсолютная оценка ошибки
ASD	Auditory Spectral Difference		Слышимая спектральная разница
Avg	Average (linear)		Среднее (линейное)
BAQ	Basic Audio Quality		Базовое качество звука
Bw	Bandwidth		Ширина полосы
CI	Confidence Interval		Доверительный интервал
CM	Coding Margin		Запас по кодированию
DBn	Database <i>n</i> (1, 2 or 3)		База данных <i>n</i> (1, 2 или 3)
DC	Direct Current		Постоянный ток
DFT	Discrete Fourier Transform	ДПФ	Дискретное преобразование Фурье
DIX	Disturbance Index		Индекс искажений
DUT	Device Under Test		Испытываемое устройство
EHS	Error Harmonic Structure		Гармоническая структура ошибок
ERB	Equivalent Rectangular Bandwidth		Эквивалентная ширина полосы частот спектра прямоугольной формы
fac	factor		Коэффициент
FFT	Fast Fourier Transform	БПФ	Быстрое преобразование Фурье
FIR	Finite Impulse Response	КИХ	Конечная импульсная характеристика
IIR	Infinite Impulse Response	БИХ	Бесконечная импульсная характеристика
ISO	International Standards Organization	ИСО	Международная организация по стандартизации
JNLD	Just Noticeable Level Difference		Едва заметная разница в уровнях

MFPD	Maximum Filtered Probability of Detection	Максимальная фильтрованная вероятность обнаружения
MOV	Model Output Variable	Выходная переменная модели
MPEG	Moving Picture Expert Group	Экспертная группа по движущимся изображениям
NL	Noise Loudness	Громкость шума
NMR	Noise-To-Mask Ratio	Отношение шума к маске
OASE	Objective Audio Signal Evaluation	Объективная оценка аудиосигнала
OCM	Objective Coding Margin	Объективный запас по кодированию
ODG	Objective Difference Grade	Объективная оценка разницы
PAQM	Perceptual Audio Quality Measure	Показатель воспринимаемого качества звука
PEAQ	Objective Measurements of Perceived Audio Quality	Объективные измерения воспринимаемого качества звука
PERCEVAL	Perceptual Evaluation	Перцептивная оценка
POM	Perceptual Objective Measure	Воспринимаемый объективный показатель
REF	Reference Signal	Эталонный сигнал
res	Resolution	Разрешение
r.m.s.	Root Mean Squared	Среднеквадратичное значение
ROEX	Rounded Exponential	Округленное экспоненциальное значение
ROV	Rate of Output Values	Скорость вывода значений
SCM	Subjective Coding Margin	Субъективный запас по кодированию
SDG	Subjective Difference Grade	Субъективная оценка разницы
SNR	Signal-to-Noise Ratio	Отношение сигнал/шум
SPL	Sound Pressure Level	Уровень звукового давления
SUT	Signal under Test	Исследуемый сигнал
THD	Total Harmonic Distortion	Полный коэффициент гармонических искажений
Win	Windowed Average	Скользящее среднее

Справочные документы

- AURAS, W. [September 1984] *Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse*. Dissertation an der Fakultät für Elektrotechnik der Technischen Universität München, Federal Republic of Germany.
- BEERENDS, J. G. and STEMERDINK, J. A. [December, 1992] A perceptual audio quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, Vol. 40, p. 963-978.
- BEERENDS, J. G. and STEMERDINK, J. A. [February, 1994] Modeling a cognitive aspect in the measurement of the quality of music codecs. Contribution to the 96th AES Convention, preprint 3800. Amsterdam, Netherlands.
- BEERENDS, J. G. and STEMERDINK, J. A. [March, 1994] A perceptual speech quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, Vol. 42, p. 115-123.
- BEERENDS, J. G., van den BRINK, W. A. C. and RODGER, B. [May, 1996] The role of informational masking and perceptual streaming in the measurement of music codec quality. Contribution to the 100th AES Convention, preprint 4176. Copenhagen, Denmark.
- BRANDENBURG, K. [1987] Evaluation of quality for audio encoding at low bit rates. Contribution to the 82nd AES Convention, preprint 2433. London, United Kingdom.
- BREGMAN, A. S. [1990] *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT Press, Cambridge MA, United States of America.
- COHEN, E. A. and FIELDER, L. D. [May, 1992] Determining noise criteria for recording environments. *J. Audio Eng. Soc.*, Vol. 40, p. 384-402.
- COLOMES, C., LEVER, M., RAULT, J. B. and DEHERY, Y. F. [April, 1995] A perceptual model applied to audio bit-rate reduction. *J. Audio Eng. Soc.*, Vol. 43, p. 233-240.
- FEITEN, B. [March, 1997] Measuring the Coding Margin of Perceptual Codecs with the Difference Signal. 102nd AES-Convention, preprint 4417. Munich, Federal Republic of Germany.
- GRUSEC, T., THIBAUT, L. and SOULODRE, G. [September, 1997] EIA/NRSC DAR systems subjective tests. Part 1: Audio codec quality. *IEEE Transactions on Broadcasting*, Vol. 43, 3.
- KARJALAINEN, J. [March, 1985] A new auditory model for the evaluation of sound quality of audio system. Proceedings of the ICASSP, p. 608-611. Tampa, Florida, United States of America.
- LEEK, M. R. and WATSON, C. S. [1984] Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, Vol. 76, p. 1037-1044.
- MEARES, D. J. and KIM, S. W. [July, 1995] "NBC time/frequency module subjective tests: overall results", ISO/IEC JTC 1/SC 29/WG 11 N0973 MPEG95/208.
- MOORE, B. C. [1986] *Frequency Selectivity in Hearing*. Academic Press, London, United Kingdom.
- MOORE, B. C. [1989] *An introduction to the psychology of hearing*. Academic Press, London, United Kingdom.
- PAILLARD, B., MABILLEAU, P., MORISSETTE, S. and SOUMAGNE, J. [1992] Perceval: Perceptual evaluation of the quality of audio signals. *J. Audio Eng. Soc.*, Vol. 40, p. 21-31.
- SCHROEDER, M. R., ATAL, B. S. and HALL, J. L. [December 1979] Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.*, Vol. 66, p. 1647-1652.
- SOULODRE, G., GRUSEC, T., LAVOIE, M. and THIBAUT, L. [March 1998] Subjective evaluation of state-of-the-art 2-channel audio codecs. *J. Audio Eng. Society*.
- SPORER, T. [October 1997] Objective audio signal evaluation – applied psychoacoustics for modeling the perceived quality of digital audio. 103rd AES-Convention, preprint 4512. New York, United States of America
- TERHARDT, E. [1979] Calculating Virtual Pitch, *Hearing Research*. Vol. 1, p. 155-182.
- THIEDE, T. and KABOT, E. [1996] A New Perceptual Quality Measure for Bit Rate Reduced Audio. Contribution to the 100th AES Convention, preprint 4280. Copenhagen, Denmark.

- TREURNIET, W. C. [1996] Simulation of individual listeners with an auditory model. Proceedings of the Audio Engineering Society, Reprint Number 4154. Copenhagen, Denmark.
- von BISMARCK, G. [1974] Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30, p. 159-172.
- ZWICKER, E. and FASTL, H. [1990] *Psycho-acoustics, Facts and Models*. Berlin; Heidelberg: Springer Verlag, Federal Republic of Germany.
- ZWICKER, E. and FELDTKELLER, R. [1967] *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel Verlag, Federal Republic of Germany.

Библиография

- GRUSEC T., THIBAUT L. and SOULODRE, G. [1995] Subjective evaluation of high quality audio coding systems: methods and results in the two-channel case. Preprint 4065 (F-5), Proceedings of the AES. New York, United States of America
-