

МСЭ-R

Сектор радиосвязи МСЭ

Рекомендация МСЭ-R BS.2132-0
(10/2019)

**Метод основанной на слышимых
различиях субъективной оценки качества
звуковых систем с использованием
нескольких входных сигналов
в отсутствие заданного
эталонного сигнала**

Серия BS
Радиовещательная служба (звуковая)



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/en>, 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/en>.)

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

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

Электронная публикация
Женева, 2020 г.

© ITU 2020

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.2132-0

Метод основанной на слышимых различиях субъективной оценки качества звуковых систем с использованием нескольких входных сигналов в отсутствие заданного эталонного сигнала

(2019)

Сфера применения

В настоящей Рекомендации описан метод основанной на слышимых различиях субъективной оценки качества звуковых систем с использованием нескольких входных сигналов в отсутствие заданного эталонного сигнала. Этот метод отражает многие аспекты метода MUSHRA, описанного в Рекомендации МСЭ-R BS.1534, но расширяет возможности оценки систем, охватывая, в отличие от Рекомендации МСЭ-R BS.1534, ситуации, когда известные скрытые эталонные и опорные сигналы отсутствуют.

Ключевые слова

Испытание с прослушиванием, качество звука, усовершенствованные звуковые системы, субъективная оценка, оценка на основе восприятия

Ассамблея радиосвязи МСЭ,

учитывая,

- a) что в Рекомендациях МСЭ-R и МСЭ-T приведено большое количество методик субъективных испытаний для оценки субъективного качества аудио-, видео- и речевых систем;
- b) что использование стандартизованных методов субъективной оценки важно для обмена данными испытаний, а также для обеспечения совместимости и корректной оценки таких данных;
- c) что для оценки усовершенствованных звуковых систем предполагается использовать стандартизованные методы испытаний;
- d) что в некоторых применениях эталонные сигналы отсутствуют или являются непригодными, так что субъективную оценку качества звуковых систем невозможно произвести путем сравнения с известным сигналом и вместо этого ее следует производить без эталонного сигнала;
- e) что в процессе производства радиовещательных программ требуется использовать технологические системы для создания звуковых сигналов и воплощения творческого замысла и что в этих случаях могут иметь место ситуации, когда целевой эталонный сигнал или эталонное поведение системы отсутствуют;
- f) что с внедрением усовершенствованных звуковых систем, описанных в Рекомендации МСЭ-R BS.2051, появляются новые инструменты для творческого выражения при производстве программ и требуются новые методы субъективной оценки, включая методы установления связи воспринимаемых атрибутов с общим воспринимаемым качеством звука,

рекомендует

для субъективной оценки звуковых систем на основе слышимых различий в отсутствие пригодного эталонного сигнала или эталонной системы использовать процедуры испытаний и оценки, приведенные в Приложении 1 к настоящей Рекомендации.

Приложение 1

1 Введение

Субъективные испытания с прослушиванием служат надежным способом измерения воспринимаемого качества аудиосистем. Существуют подробно описанные и апробированные методы оценки качества звука радиовещательных передач, когда системы сравниваются с известным неискаженным эталонным сигналом как при высоком, так и при среднем уровнях качества. Рекомендация МСЭ-R BS.1116 "Методы субъективной оценки небольшого ухудшения качества в звуковых системах", включая многоканальные звуковые системы, предназначена для оценки высококачественных звуковых систем с небольшим ухудшением качества по сравнению с заданным эталонным сигналом, а Рекомендация МСЭ-R BS.1534 "Метод субъективной оценки промежуточного уровня качества аудиосистем" – для оценки аудиосистем среднего уровня, подходящих для широковещательных применений, но демонстрирующих явное отличие от эталонного сигнала. Следует отметить, что разработка этих двух методов была в значительной степени обусловлена необходимостью оценки влияния систем кодирования с низкой скоростью передачи.

В некоторых применениях эталонный сигнал отсутствует или является непригодным, и субъективное качество систем не может быть оценено на соответствие эталону. В Рекомендации МСЭ-R BS.1284 "Общие методы субъективной оценки качества звука" описаны только методы, которые либо ориентированы на высококачественный звук, либо не предусматривают абсолютной шкалы качества звука.

В настоящей Рекомендации описан метод субъективной оценки качества звуковых систем на основе слышимых различий с использованием нескольких входных сигналов в применениях, когда эталонный сигнал отсутствует.

При этом методе в качестве основы для сравнения испытуемых звуковых систем используется подход представления нескольких входных сигналов, изложенный в Рекомендации МСЭ-R BS.1534. Оценщику предлагают дать оценку каждой испытуемой системе по следующим критериям:

- 1) общее субъективное качество звука;
- 2) оценка атрибутов (заранее определенных наборов избранных атрибутов).

Оценка общего субъективного качества звука производится с использованием непрерывной шкалы качества, определение которой приведено в Рекомендации МСЭ-R BS.1534.

В силу того, что этап оценки атрибутов, описанный в Рекомендациях МСЭ-R BS.1116, МСЭ-R BS.1284, МСЭ-T P.835 и МСЭ-T P.806, не обязателен, для каждого эксперимента предварительно подбираются несколько подходящих атрибутов качества звука, предпочтительно из существующих и проверенных лексиконов. Оценщики оценивают эти атрибуты по 100-балльной линейной шкале.

С помощью статистического анализа этих двух типов данных можно сделать следующие выводы:

- относительная общая субъективная оценка качества звука каждой звуковой системы;
- воспринимаемые характеристики избранных атрибутов каждой звуковой системы (факультативно);
- относительный вес различных воспринимаемых характеристик в общей оценке воспринимаемого качества испытуемых систем (факультативно).

2 Терминология

Общее субъективное качество – единый атрибут, охватывающий все аспекты оцениваемого качества звука, то есть базовое качество звука, как определено в Рекомендации МСЭ-R BS.1284. Во избежание возможной путаницы с термином "базовое качество звука", определенным в Рекомендации МСЭ-R BS.1116, здесь используется термин "общее субъективное качество".

Основное различие между базовым качеством звука и общим субъективным качеством заключается не в различии воспринимаемых атрибутов, которые остаются общими для этих двух всеохватывающих

терминов оценки качества, а в различии процессов его оценки. Оценка базового качества звука выполняется путем сравнения двух или более входных сигналов, причем один из них определяется как эталонный (например, в целях определения количественной разницы между сжатой версией звукового элемента и несжатым оригиналом). Напротив, общее субъективное качество – это количественная оценка разницы только с внутренним эталоном – то есть с ожиданием слушателя – при отсутствии внешнего эталона (например, разных бинауральных представлений).

Управляемые переменные – это переменные, которыми управляют в ходе эксперимента, что позволяет структурировать и контролировать его конфигурацию. Их еще называют независимыми переменными, потому что их значения не зависят от других экспериментальных переменных.

Ответные переменные – переменные, которым оценщики дают оценку, когда оценивают воспринимаемые сигналы по заданной шкале. Их еще называют зависимыми переменными, поскольку их значения зависят от других экспериментальных переменных, то есть от независимых/управляемых переменных.

Условия – набор значений управляемых переменных, которые используются при оценке.

Испытание – этап процесса оценки, на котором оцениваемые системы (или их подмножество) представляются при заданных условиях, и оценщик дает свою оценку в виде ответной переменной.

Повторная оценка – повторение условий испытания, когда одни и те же ответные переменные оцениваются оценщиком при одинаковых значениях управляемых переменных.

Дескриптивный – основанный на объективном описании без высказывания суждения.

Атрибут – определенная характеристика воспринимаемого качества, которую можно оценить с использованием шкалы оценок. Общее воспринимаемое субъективное качество может оцениваться с учетом множества атрибутов.

Программный элемент – фрагмент аудиоматериала, используемый в ходе оценки в сочетании с другими управляемыми переменными.

Входной сигнал – индивидуальное представление элемента программы через систему при определенном наборе значений управляемых переменных.

Лексикон – набор дескриптивных воспринимаемых атрибутов с четкими именами, определениями и шкалами оценки.

3 Общепринятая практика

Для сбора надежной информации, представляющей научный интерес, применяются различные стратегии проведения исследований. При проведении субъективной оценки качества аудиосистем необходимо применять формальные экспериментальные методы, способные гарантировать достоверность результатов и их интерпретации. Для того чтобы в ходе субъективных экспериментов получить надежные данные, необходимо управлять их условиями и изменять их, поскольку эксперимент, представляемый оценщикам с возможностью управления, дает высококачественные результаты. Необходимо тщательно проектировать и планировать эксперимент, чтобы минимизировать неконтролируемые факторы, которые могут добавить нежелательный шум. Например, если все условия эксперимента представляются всем оценщикам в виде одной и той же фиксированной последовательности, то это может привести к систематической ошибке, обусловленной влиянием порядка, которую невозможно исключить из данных и их интерпретации. Наилучшей практикой в этом отношении является предоставление условий каждому оценщику в случайном порядке или с использованием сбалансированной схемы, с тем чтобы свести к минимуму любые потенциальные систематические ошибки, обусловленные влиянием порядка. Рекомендуемая процедура испытания подробно описана в разделе 5.

Для того чтобы гарантировать наилучшее качество данных в таких экспериментах, важно принять во внимание некоторые из следующих положений, входящих в состав настоящей Рекомендации.

Должны быть привлечены опытные оценщики, так как они обычно выдают высококачественные данные. Опытные оценщики отбираются и проверяются, как описано в Рекомендации МСЭ-R BS.1534. Для того чтобы иметь возможность корректно оценить характеристики испытуемых систем, также важно использовать критичные программные элементы и выбирать такие воспринимаемые атрибуты, которые наилучшим образом дифференцируют испытуемые системы и способствуют оценке воспринимаемого качества.

Что касается проектирования эксперимента, то экспериментатор должен тщательно спланировать продолжительность периода его проведения. Включение достаточного количества и диапазона критичных испытательных элементов позволит получить в большей степени поддающееся обобщению представление о рабочих характеристиках исследуемых звуковых систем. Также желательно сравнивать большое количество звуковых систем. Это обычные цели, но с ними связаны дополнительные затраты времени и средств, а также риск перегрузки оценщиков. Методы, облегчающие планирование ресурсов (включая оценку продолжительности испытаний), описаны в информативном Прилагаемом документе 1 к Приложению 1 к настоящей Рекомендации.

Для того чтобы эксперимент можно было точно продолжить или воспроизвести позднее или в другом месте, в протокол испытаний необходимо включить не только его результаты, но и все детали их проведения. Руководящие указания по составлению протоколов изложены в Рекомендациях МСЭ-R BS.1116 и МСЭ-R BS.1534.

4 Параметры эксперимента

В данном разделе определены ключевые параметры, позволяющие составить структурированную конфигурацию управляемого эксперимента. Они делятся на две основные категории, а именно: управляемые переменные эксперимента и ответные переменные.

4.1 Управляемые переменные эксперимента

Управляемые переменные эксперимента (или независимые переменные) используются для определения параметров, которыми управляют в ходе эксперимента, что позволяет получить структурированную и управляемую конфигурацию эксперимента, способствующую проведению тщательного статистического анализа. Как правило, управляемые переменные определяются для таких параметров, как оцениваемые системы, программный материал, оценщики и повторные оценки. Для каждой управляемой переменной экспериментатор определяет количество уровней. Например, включение в испытание 10 разных программных элементов соответствует наличию 10 уровней переменной "программные элементы". Затем это количество уровней используется при планировании эксперимента и последующем статистическом анализе.

4.1.1 Оцениваемые системы

В таких экспериментах экспериментатор занимается изучением воспринимаемого качества оцениваемой технологии или системы.

Для минимизации погрешности оценки количество оцениваемых систем в соответствии с законом Миллера (Miller, G.A., 1956) должно находиться в диапазоне от 5 до 9. Если желаемое количество звуковых систем для оценки превышает 9, более подробные указания можно найти в пункте 5.1.

По возможности экспериментатору следует включить в эксперимент одну или несколько систем с известным качеством, чтобы результаты по испытуемым системам можно было рассматривать в контексте.

4.1.2 Программный материал

Выбор исследуемого материала должен соответствовать процедурам, изложенным в Рекомендациях МСЭ-R BS.1116 и МСЭ-R BS.1534. Поскольку универсального программного материала, который можно было бы использовать для оценки любых систем при любых условиях, не существует, критичный программный материал следует специально изыскивать. Как правило, поиск хорошего материала – трудоемкая работа; но если не найти критичный материал для каждой системы, эксперименты не выявят различий между системами и будут неубедительными.

4.1.3 Оценщики

Для обеспечения надлежащего качества собираемых экспериментальных данных рекомендуется привлекать опытных оценщиков. Такие оценщики должны иметь опыт критического прослушивания звуковых материалов. Они дадут более надежные результаты, чем неопытные оценщики, затратив на это меньше времени. Также важно отметить, что большинство неопытных оценщиков после частого прослушивания обычно становятся чувствительнее к разнообразным артефактам. Опытного оценщика выбирают по его способности выполнить испытание на прослушивание. Эта способность должна быть качественно и количественно охарактеризована путем испытания с повторным оцениванием по параметрам надежности и различительной способности, как указано ниже:

- 1) различительная способность – мера способности воспринимать различия между испытательными элементами;
- 2) надежность – мера близости повторных оценок одного и того же испытательного элемента.

К окончательному анализу данных следует привлекать только оценщиков, признанных опытными по каждому проведенному испытанию (см. Рекомендацию МСЭ-R BS.1116). Основанием для этого служит как минимум одна повторная оценка от каждого оценщика, что позволяет качественно и количественно охарактеризовать уровень квалификации оценщика в ходе одного эксперимента. Такие методы применяются либо на этапе первичного отбора оценщиков в рамках предварительного эксперимента, либо, что предпочтительнее, как на этапе первичного отбора, так и в ходе последующего отсеивания (с использованием ответов, полученных в ходе основной оценки). Предварительный эксперимент часто представляет собой оценку меньшего масштаба, связанную с основным экспериментом, и охватывает репрезентативную подборку образцов, подлежащих оценке в основном эксперименте. Для оценки квалификации слушателя в предварительном эксперименте следует использовать релевантное подмножество испытательных входных сигналов, адекватно отражающих все множество входных сигналов и артефактов, оцениваемых в ходе основного эксперимента.

Как правило, должны быть задействованы около 20 (предпочтительно больше) опытных оценщиков.

4.1.4 Повторная оценка

Для оценки качества собранных данных и эффективности работы оценщиков каждому оценщику предлагают дать повторную оценку условий. Предполагается, что для проверки эффективности работы оценщика без чрезмерного увеличения размера эксперимента следует повторно оценить по крайней мере два образца (см. предыдущий пункт).

4.1.5 Дополнительные управляемые переменные эксперимента

Могут быть случаи, когда для эксперимента требуются дополнительные управляемые переменные. Это вполне нормально и приемлемо, и эти переменные можно аналогичным структурированным образом добавить к тем, что определены в пункте 4.1. Экспериментатору следует знать, что увеличение количества управляемых переменных ведет к увеличению размера и продолжительности эксперимента.

4.1.6 Ответные переменные

По каждому условию оценщикам предлагают дать свою оценку, используя ответные переменные. Следует использовать ответные переменные двух разных типов, описанных ниже, с указанием их размерности:

- общее субъективное качество (по каждой системе);
- оценки атрибутов (факультативно, по каждой системе).

4.1.7 Общее субъективное качество звука

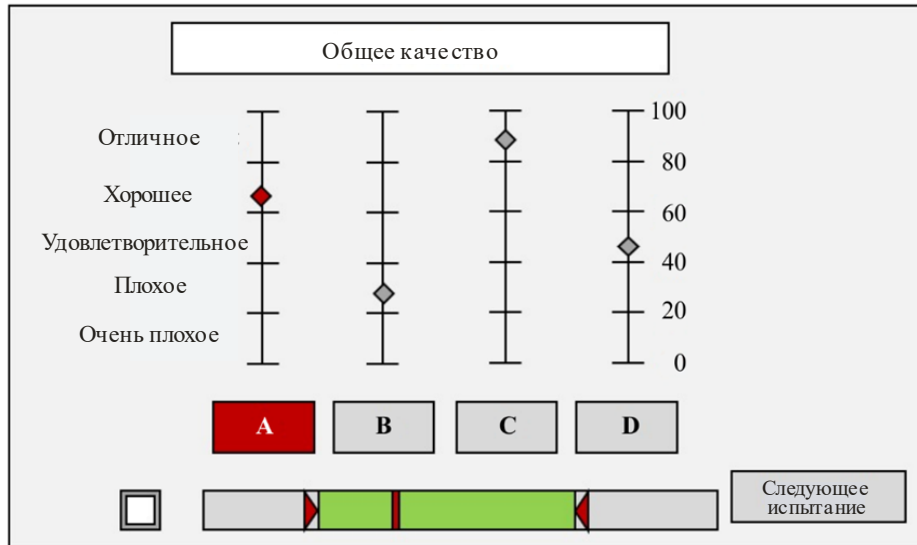


Сначала оценщикам предлагается дать общую субъективную оценку качества звука с использованием непрерывной шкалы качества¹ (CQS). Оценщиков просят дать общее субъективное суждение о качестве звука каждого образца и представить свою оценку по CQS. CQS – это 100-балльная линейная шкала (обычно > 10 см), разделенная на пять равных интервалов с определениями, как показано на рисунке 1. В единичном испытании с общим программным элементом представляется несколько систем, у каждой из которых своя собственная шкала оценок, как показано на рисунке 2.

¹ Эта шкала также используется для оценки качества изображения (Рекомендация МСЭ-R BT.500 "Методика субъективной оценки качества телевизионных изображений" и Рекомендация МСЭ-R BS.1534 "Метод субъективной оценки промежуточного уровня качества аудиосистем").

РИСУНОК 2

Пример графического интерфейса пользователя для общей субъективной оценки качества звука



BS.2132-02

4.1.8 Оценки атрибутов

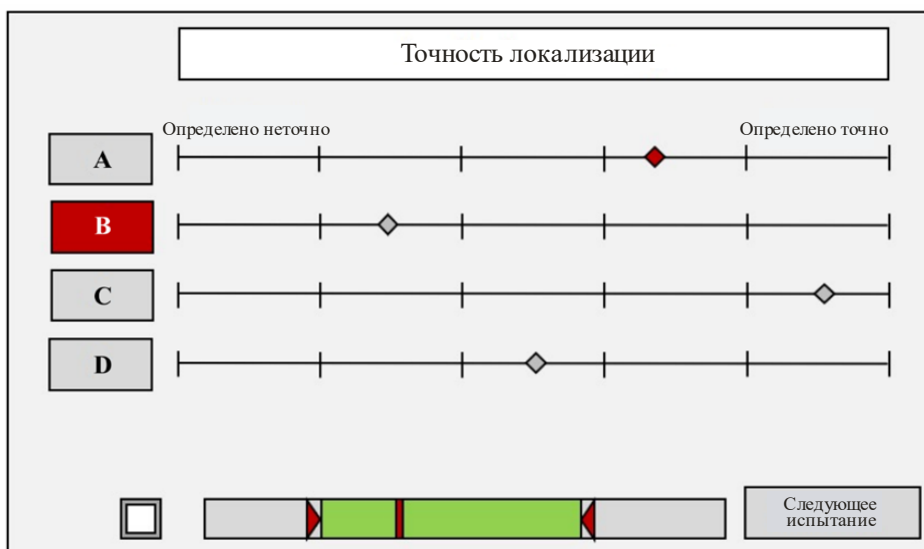
После проведения общей субъективной оценки качества звука оценщикам дополнительно предлагается оценить комбинацию система – программный элемент по каждому из дескриптивных атрибутов. И вновь в единичном испытании с общим программным элементом представляется несколько систем, у каждой из которых своя собственная шкала оценок.

Как правило, используются непрерывные 100-балльные шкалы.

На рисунке 3 показан пример интерфейса для оценки параметра, когда параметром и его определением является точность локализации.

РИСУНОК 3

Пример графического интерфейса пользователя для оценки атрибута (*точность локализации*)



BS.2132-03

5 Протокол испытаний

5.1 Планирование эксперимента

Для того чтобы гарантировать получение высококачественных данных в случае минимизации источников случайных или неконтролируемых эффектов, эксперимент необходимо тщательно спланировать. Планирование полезно и для оценки масштаба и продолжительности эксперимента, а также для получения конфигурации для статистического анализа. План включает два ключевых аспекта, а именно: план обработки и план распределения входных сигналов, как описано ниже.

5.1.1 План обработки

План обработки определяет, какие управляемые переменные следует использовать в эксперименте, за исключением (переменной оценщика).

Для экспериментов среднего размера рекомендуется полный факторный план, когда оцениваются все возможные комбинации уровней управляемых переменных. Количество условий для эксперимента с полным факторным планом получают путем перемножения количества уровней по каждой из независимых переменных.

5.1.2 План распределения входных сигналов

План распределения входных сигналов определяет, каким образом условия предоставляются каждому оценщику.

Рекомендуется внутрисубъектный план, при котором каждому оценщику для оценки предоставляются все условия. Для того чтобы минимизировать систематические ошибки смещения, порядок представления регулируется, как правило, путем рандомизации. Желательно обеспечить полностью сбалансированный порядок представления.

5.1.3 Разбиение крупных экспериментов

В определенных ситуациях при использовании полного факторного внутрисубъектного плана эксперимент может оказаться слишком крупным и громоздким. Это случается, например, когда необходимо оценить большое количество звуковых систем или когда общая продолжительность испытания для каждого оценщика становится неоправданно большой.

В таких случаях стоит рассмотреть возможность усовершенствованного плана эксперимента.

В данном разделе лишь иллюстрируются подходы, которые можно рассмотреть в таких случаях. Заинтересованному экспериментатору следует ознакомиться с литературой по планированию экспериментов, где можно найти соответствующее руководство.

Ниже в качестве примеров приведены два разных решения для проведения крупных экспериментов.

5.1.3.1 План с расщепленными делянками

В пункте 4.1.1 рекомендуется ограничить максимальное количество оцениваемых систем 5–9 системами. Если нужно оценить большее количество звуковых систем, можно рассмотреть возможность использования плана с расщепленными делянками. Например, если требуется оценить 14 звуковых систем, полная оценка может состоять из двух испытаний, включающих по семь звуковых систем каждое. Для того чтобы контролировать любые эффекты смещения, связанные с отдельным представлением, распределение звуковых систем по испытаниям следует производить случайным образом. Однако подобная рандомизация не должна влиять на общее количество условий, предоставляемых каждому оценщику или в рамках всего эксперимента. См. пункт 5.1.1.

При использовании плана с расщепленными делянками важно, чтобы в анализ были включены факторы, связанные с объединением в блоки. В идеале объединение в блоки не должно являться статистически значимым фактором.

5.1.3.2 Межсубъектный план

Если количество условий, подлежащих оценке одним оценщиком, велико, это может привести к неоправданно большому количеству сеансов прослушивания, особенно когда общая продолжительность испытаний может превысить четыре часа на каждого оценщика.

Один из подходов к решению этой проблемы – использование межсубъектного (или межгруппового) плана. Это общий метод, при котором разным оценщикам или группам оценщиков предлагаются разные условия. Простой способ уменьшить количество условий, предлагаемых каждому оценщику (или группе оценщиков), состоит в том, чтобы каждому оценщику (или группе) выделить свои поднаборы программных элементов. Это следует делать с осторожностью и обеспечить, чтобы общее количество предлагаемых условий было пропорционально распределено между всеми оценщиками (или группами оценщиков).

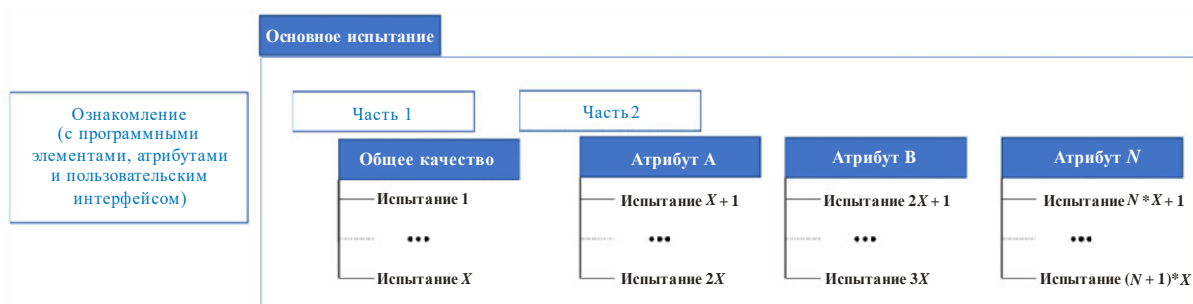
При использовании такого плана важно, чтобы факторы группирования были включены в анализ. В идеале группирование не должно являться статистически значимым фактором.

5.2 Проведение испытания

Испытание следует проводить в два этапа плюс ознакомление, за которым следует основное испытание с использованием атрибутов. Порядок, в котором оценщикам представляются различные элементы, показан на рисунке 4. Предполагается, что оцениваются N атрибутов с использованием M программных элементов. Всего проводится x испытаний для общей субъективной оценки качества звука и y испытаний для оценки атрибутов.

РИСУНОК 4

Блок-схема общей конфигурации испытания, включая ознакомление и основное испытание



BS.2132-04

Для того чтобы обеспечить сбор высококачественных данных, оценщиков следует ознакомить с протоколом испытаний, пользовательскими интерфейсами, программными элементами, а также с воспринимаемыми атрибутами. Кроме того, каждому оценщику должна быть дана возможность прослушать испытательные входные сигналы, ознакомиться с атрибутами и опробовать пользовательский интерфейс. Сюда можно отнести произвольное и слепое прослушивание входных сигналов из поднабора испытаний, производимое исключительно в целях ознакомления. Описанный выше предпочтительный процесс показан на рисунке 5. В процессе испытаний оценщик оценивает отдельные входные сигналы для отдельного атрибута, используя индивидуальные интерфейсы.

РИСУНОК 5

Процесс оценки индивидуальных входных сигналов с использованием индивидуальных интерфейсов



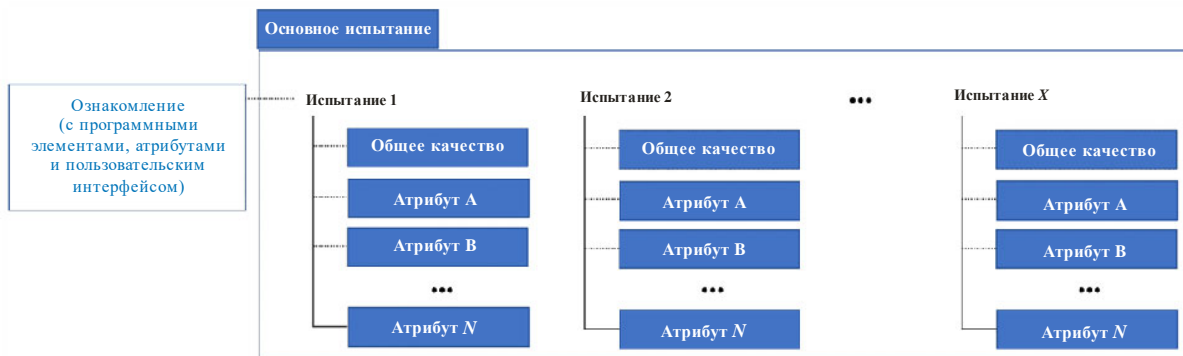
BS.2132-05

5.2.1 Факультативная процедура

На рисунке 6 представлена факультативная часть общего испытания. Оценщику могут быть представлены отдельные интерфейсы для каждого атрибута или составной интерфейс для нескольких атрибутов. На рисунке 7 представлен пример графического интерфейса пользователя (GUI), в котором оценщик оценивает каждый входной сигнал, используя отдельные интерфейсы для каждого атрибута. На рисунке 8 представлен пример GUI, когда оценщик оценивает несколько атрибутов с использованием составного интерфейса.

РИСУНОК 6

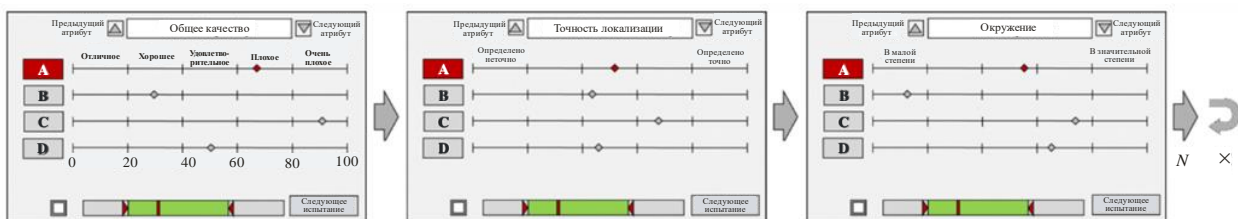
Факультативная блок-схема общей конфигурации испытаний, включая ознакомление и основное испытание



BS.2132-06

РИСУНОК 7

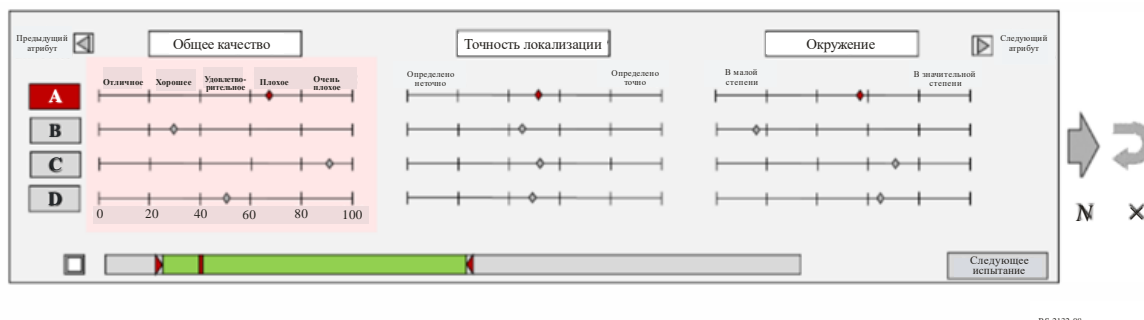
Процесс оценки каждого атрибута одного и того же входного сигнала с использованием индивидуальных интерфейсов



BS.2133-07

РИСУНОК 8

Процесс оценки нескольких параметров одного и того же входного сигнала с использованием составного интерфейса



5.3 Инструктаж оценщиков

Для того чтобы гарантировать, что оценщики полностью осведомлены о задаче, которую предстоит выполнить, до начала эксперимента им следует выдать инструкции – как письменные, так и устные. Инструкции должны проинформировать оценщиков о задаче, которую им предстоит выполнить без чрезмерных систематических ошибок, и ознакомить их с протоколом испытаний, ответными переменными (общее субъективное качество звука, определения атрибутов) и порядком их использования с помощью испытательного GUI. Примеры таких инструкций приведены в Прилагаемом документе 2.

6 Условия проведения испытаний

6.1 Условия прослушивания

Испытания должны проводиться в помещении для прослушивания, соответствующем Рекомендации МСЭ-R BS.1116.

6.2 Воспроизводящие устройства

Для испытаний можно использовать наушники или громкоговорители, однако использование того и другого в ходе одного сеанса испытаний не допускается. Все оценщики должны использовать преобразователь одинакового типа. Следует использовать эталонные контрольные громкоговорители или наушники, как указано в Рекомендации МСЭ-R BS.1116.

Предпочтительно, чтобы конфигурация громкоговорителей и предъявляемые к ним требования, а также позиции прослушивания, занимаемые оценщиками, выбирались в соответствии с требованиями, изложенными в Рекомендации МСЭ-R BS.1116.

Предпочтительно, чтобы схемы расположения громкоговорителей выбирались из конфигураций, определенных в Рекомендации МСЭ R BS.2051.

6.3 Калибровка

Для того чтобы гарантировать повторяемые и воспроизводимые результаты, важно тщательно настроить оборудование и материалы, которые будут использоваться для испытаний.

Относительная громкость элементов

Относительную громкость различных элементов не следует как-либо связывать с измерениями громкости, описанными в Рекомендации МСЭ-R BS.1770. Короткие звуковые фрагменты должны быть настроены на запланированный уровень громкости. Следует сохранять разницу между громким (фортиссимо) и тихим (пианиссимо) элементами, которая может составлять, например, 15 дБ. Относительная громкость каждого элемента должна регулироваться субъективно, предпочтительно группой опытных оценщиков. Для того чтобы поддерживать эту разницу на уровне воспроизведения, важно надлежащим образом испытать разные элементы.

Относительная громкость входных сигналов

Небольшие различия в громкости обычно вносят систематическую погрешность в пользу более громкого сигнала. Такие различия между входными сигналами одного и того же элемента должны быть устранены. Следует использовать объективное (а не субъективное) измерение громкости, как указано в Рекомендации МСЭ-R BS.1770. Если объективные измерения невозможны, громкость каждого фрагмента перед его включением в испытание должна быть субъективно скорректирована группой квалифицированных оценщиков.

Синхронизация элементов

Мгновенное переключение между входными сигналами, представляющими собой по-разному обработанные версии одного и того же программного элемента, не должно приводить к ощутимому временному сдвигу. Подробные инструкции по представлению входных сигналов приведены в Рекомендации МСЭ-R BS.1534.

6.3.1 Калибровка системы воспроизведения

При проведении испытаний с использованием громкоговорителей предпочтительно применять одну из схем их расположения, указанных в Рекомендациях МСЭ-R BS.775 и МСЭ-R BS.2051. В ином случае для описания схемы расположения, применяемой при испытании, следует использовать систему обозначений, указанную в соответствующей Рекомендации.

Уровень системы воспроизведения должен быть отрегулирован, как описано в Рекомендации МСЭ-R BS.1116.

Подробные сведения о калибровке сабвуферов и систем управления низкими звуковыми частотами выходят за рамки настоящего документа. Результатом управления низкими частотами должна стать горизонтальная (в пределах допуска, указанного в Рекомендации МСЭ-R BS.1116) частотная характеристика отдельной комбинации громкоговорителей и сабвуфера.

В ходе предыдущих испытательных последовательностей было замечено, что отдельные слушатели могут предпочитать разные абсолютные уровни прослушивания. Хотя это и не приветствуется, не всегда удается воспрепятствовать требованиям обеспечения такой степени гибкости, выдвигаемым участниками. В настоящее время нет точных данных о том, влияет ли подобная гибкость на слышимость некоторых оцениваемых артефактов. Поэтому если участники регулируют усиление системы, этот факт следует отразить в результатах испытаний.

Разница во времени задержки между каналами стереофонической системы не должна превышать 20 мкс для наушников и 100 мкс для громкоговорителей, как указано в Рекомендации МСЭ-R BS.1116.

При использовании систем с сопровождающим изображением совокупное время задержки эталонного контрольного громкоговорителя и исследуемой(ых) системы (систем) не должно превышать пределов, установленных в Рекомендации МСЭ-R BS.775.

ПРИМЕЧАНИЕ. – Измерение акустических параметров усовершенствованных звуковых систем может оказаться гораздо сложнее, чем в случае более ранних многоканальных звуковых систем. При осуществлении измерений следует внимательно подходить к выбору измерительного микрофона и его ориентации, см. Отчет МСЭ-R BS.2419. Дополнительная информация об электроакустических требованиях и акустических характеристиках рабочего помещения приведена в Рекомендации МСЭ-R BS.1116.

7 Статистический анализ

Для обеспечения более глубокого понимания субъективного качества систем проводится статистический анализ данных, полученных от оценщиков. Рассчитываются характеризующие его средние оценки и проводится оценка дисперсии данных, которая используется для определения достоверности наблюдаемых различий в полученных характеристиках систем.

При сборе данных каждый оценщик дает оценку атрибутам оцениваемых систем. Эти системы испытываются с разными программными элементами. Оценщик оценивает каждую систему по списку воспринимаемых атрибутов, используя предварительно определенные шкалы. Для каждого программного элемента каждый оценщик оценивает каждый атрибут, используя один и тот же набор

шкал атрибутов. Оценщик также оценивает каждую комбинацию система – программный элемент в отношении общего субъективного качества.

Для каждого программного элемента оценщики должны дать свою оценку атрибутов каждой системы, а также общую субъективную оценку качества звука.

В Рекомендации МСЭ-R BS.1534 приведены детали статистического анализа данных по общему субъективному качеству звука и по каждому дескриптивному атрибуту.

Кроме того, анализ данных по восприятию, полученных при использовании методов, приведенных в настоящей Рекомендации, дает сопоставимые результаты с анализом данных по восприятию, полученных с использованием более классического количественного дескриптивного анализа. Такой анализ включают в себя дисперсионный анализ (ANOVA), выполняемый для каждого воспринимаемого атрибута, а также многофакторный анализ (например, анализ главных компонент, PCA).

8 Библиография

- [1] Рекомендация МСЭ-R BS.2051 – Усовершенствованная звуковая система для производства программ
- [2] Рекомендация МСЭ-R BS.775 – Многоканальные стереофонические звуковые системы с сопровождающим изображением и без него
- [3] Recommendation ITU-R BS.645 – Test signals and metering to be used on international sound programme connections
- [4] Рекомендация МСЭ-R BS.1116 – Методы субъективной оценки небольшого ухудшения качества в звуковых системах
- [5] Рекомендация МСЭ-R BS.1534 – Метод субъективной оценки промежуточного уровня качества аудиосистем
- [6] Рекомендация МСЭ-R BS.1770 – Алгоритмы измерения громкости звуковых программ и истинного пикового уровня звукового сигнала
- [7] Рекомендация МСЭ-R BS.1864 – Эксплуатационная практика в отношении громкости при международном обмене программами цифрового телевидения
- [8] Miller, G.A. (1956), The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63 (2): 81–9
- [9] Рекомендация МСЭ-R BT.500-13 – Методика субъективной оценки качества телевизионных изображений
- [10] Recommendation ITU-T P.835 – Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm
- [11] Recommendation ITU-T P.806 – A subjective quality test methodology using multiple rating scales
- [12] Report ITU-R BS.2419 – Effect of microphone directivity regarding level calibration and equalization of advanced sound systems

Прилагаемый документ 1 к Приложению 1 (информационный)

Инструмент Excel для оценки испытательного периода эксперимента

Предоставленный ниже инструмент Excel предназначен для оценки испытательного периода эксперимента по субъективной оценке звука в целях планирования ресурсов. Пример эксперимента приведен в таблице 1. Светло-коричневые поля представляют собой поля ввода данных. Светло-зеленые – поля выходных данных.



Copy of
R15-WP6C-190715-T

ТАБЛИЦА 1

Экранный снимок приведенной таблицы Excel для оценки испытательного периода эксперимента по субъективной оценке звука

Full Factorial Design Calculator						Input fields			
v4						Result fields			
Controlled experimental variables						Test parameters			
(Independent variables)									
Variable	Label	Description	No. of levels	calculation no of levels	Degrees of Freedom (DOF)	Parameter	No.	Units	Comments
1	SYSYEM	No. of systems under test	7	7	6	No. of controlled experimental conditions (total)	21		Excluding assessors
2	PROGRAMME	No. of programme items	3	3	2	No. of test conditions (per replication)	21		Excluding assessors
3	REPLICATE	No. of presentations or repetitions	1	1	0	Total no. of independent variables	420		Including assessors
4	ASSESSOR	No. of assessors	20	20	19				
Total			31		27	No. of blocks	1		1 = within-subjects design >1 = between-subjects design
Response variables						No. of PROGRAMME items per block	3		Must be integer ≥2
(Dependent variables)						No. of ratings per condition	20		
Variable	Label	Description	No. of levels			Total no. of ratings per assessor	147		
Total		Overall subjective Quality	1			Estimated average rating time per condition	20	s	
Descriptive		Envelopment, source width, etc.	6			Estimated total duration per assessor	0,8	hrs	
						Session duration	2	hrs	max. < 2 hrs incl. breaks
						Estimates no. of sessions per assessor	1	Sessions	
						Total no. of sessions	20	Sessions	
						No. of data points per response variable	420		
						Total no. of data points in experiment	2940		

Прилагаемый документ 2 к Приложению 1 (информационный)

Пример инструкций для оценщика

A2.1 Общие инструкции

Уважаемый слушатель!

Добро пожаловать в команду участников данного эксперимента, в ходе которого вам предстоит прослушать и оценить различные звуковые системы с разными программными элементами. При испытаниях используется метод "несколько входных сигналов в отсутствие заданного эталонного сигнала".

Испытание, продолжающееся в общей сложности около двух часов, состоит из ознакомительного этапа, за которым следует основной этап. В процессе ознакомления вы получите сведения о программных элементах, пользовательских интерфейсах, а также атрибутах, используемых для оценки каждой звуковой системы. После ознакомительного этапа вы должны будете выполнить испытание, состоящее из двух частей.

Часть 1 заключается в оценке звуковых систем в отношении *общего субъективного качества*.

Общее субъективное качество – это единый атрибут, охватывающий все аспекты оцениваемого качества звука.

Часть 2 заключается в оценке каждой звуковой системы в отношении следующих атрибутов:

- глубина сцены;
- окружение;
- погружение;
- точность локализации;
- яркость;
- искажение.

Определения этих атрибутов приведены ниже и перед началом испытания будут объяснены.

При выполнении каждого этапа испытания внимательно прослушайте программные элементы, заметив время, необходимое для изучения и оценки каждого из них.

Если у вас есть какие-либо вопросы или вам требуются дополнительные разъяснения по протоколу испытаний, задавайте их, желательно на этапе ознакомления или после него.

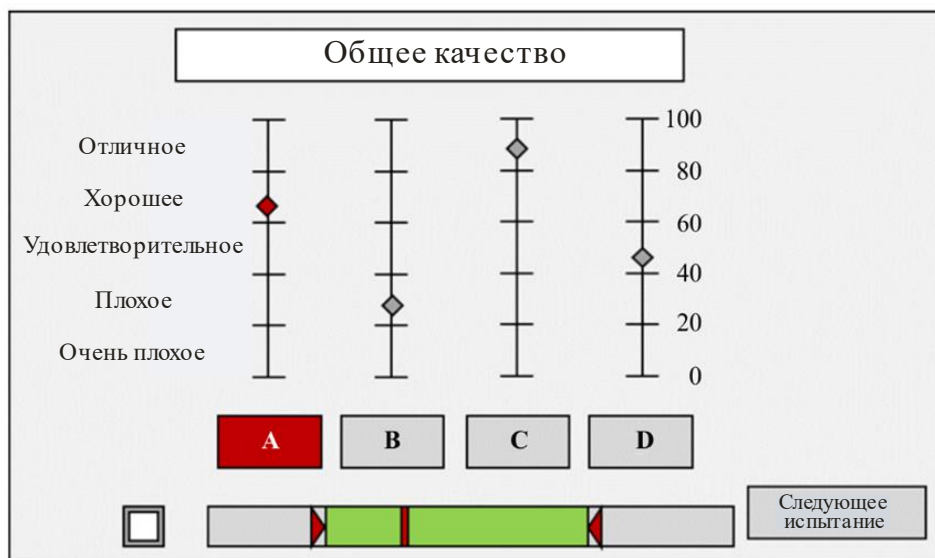
Рекомендуется делать перерыв через каждые 20–30 минут, чтобы размять ноги и немного отдохнуть.

A2.1.1 Общая оценка качества

Вам будет предложено дать общую оценку качества каждого звукового фрагмента в отношении *общего субъективного качества*, используя непрерывную 100-балльную шкалу качества, показанную на рисунке 9. Внимательно прослушайте каждый фрагмент необходимое число раз и дайте свою оценку, когда будете готовы. Дав оценку всех фрагментов, нажмите кнопку "Следующее", чтобы перейти к следующему испытанию.

РИСУНОК 9

Пользовательский интерфейс для общей субъективной оценки качества звука



BS.2132-09

A2.1.2 Дескриптивные атрибуты

В ходе каждого испытания вам предложат оценить качество звука каждой из систем по одному из атрибутов (см. таблицу 2).

ТАБЛИЦА 2

Отметки и определения атрибутов

Атрибут	Определение	Пример	Нижняя отметка	Верхняя отметка
Глубина сцены	Воспринимаемая глубина звукового образа. Учитывает как общую глубину сцены, так и относительное расстояние до отдельных источников звука		Плоская	Глубокая
Окружение	Создает ли воспроизводимый звук впечатление, что вы окружены им, и ощущение окружающего или охватывающего вас пространства? Ощущение, что вы окружены звуком	В какой степени звук	В малой степени	В значительной степени
Погружение	Воспринимаемая высота расположения источника звука. Ощущение захлестывания до полного окружения или покрытия	Насколько источник звука кажется расположенным выше или ниже вас? В какой степени звук окружает вас по вертикали	В малой степени	В значительной степени
Точность локализации	Насколько правильно отдельные инструменты и голоса размещены и разделены в пространственном звуковом образе? Насколько точно отдельные источники звука локализируются в помещении? Если отдельные источники звука случайным образом разнесены слишком широко, точность локализации будет низкой. Можно ли точно разместить и разделить отдельные инструменты и голоса в пространственном звуковом образе? Насколько точно отдельные источники звука позиционированы в помещении?		Определяется неточно	Определяется точно

ТАБЛИЦА 2 (окончание)

Атрибут	Определение	Пример	Нижняя отметка	Верхняя отметка
Яркость	Высокочастотное расширение: – незначительное – музыка слышна как будто через дверь, приглушенно, размыто или тускло – значительное – легкость, чистота и ясность с пространством для инструментов. Четкие верхние частоты без резкости и пронзительности и без искажений		Незначительное	Значительное
Искажение	Дополнительные и нежелательные звуки, добавляющие к звуку артефакты	Может проявляться в форме временных или тембральных артефактов, приводящих, например, к резкому, скрипучему или надтреснутому звуку или к звону	Незначительное	Значительное

РИСУНОК 10

Пользовательский интерфейс для оценки дескриптивных атрибутов

Точность локализации

Определено неточно | Определено точно

A | B | C | D

Следующее испытание

BS.2132-10

Приступайте к ознакомлению, когда будете готовы.

Благодарим за участие!

**Прилагаемый документ 3
к Приложению 1
(информационный)**

**Примеры субъективной оценки и дескриптивного профилирования
качества звука аудиосистем в отсутствие заданного эталонного сигнала**

Внедрение усовершенствованных звуковых систем, описанных в Рекомендации МСЭ-R BS.2051, обеспечивает инструменты для творческого самовыражения при производстве программ. При оценке этих систем могут возникать условия, когда известные скрытые эталонные и опорные сигналы отсутствуют. В этих условиях применяются методы, описанные в настоящей Рекомендации.

Возможны и другие примеры, когда экспериментаторы могут воспользоваться этой методикой для характеристики субъективного качества своих систем или сигналов.

Примерами применения этого метода могут служить субъективные оценки качества по следующим пунктам:

- поведение производственных рендереров усовершенствованных звуковых систем в отсутствие или при непригодности эталонного сигнала, подходящего для планов продюсера;
 - воспроизведение сложной звуковой программы при разных схемах расположения громкоговорителей с помощью одного производственного рендерера;
 - системы для воспроизведения контента усовершенствованных звуковых систем в домашнем кинотеатре в отсутствие системы, которая задавала бы априори известный эталон качества;
 - алгоритмы повышающего или понижающего микширования;
 - система направленных микрофонов для пространственной аудиозаписи и производства программ;
 - процессоры реверберации для производства пространственного звука;
 - системы обработки сигналов для многополосных громкоговорителей и радиовещательных установок;
 - методы восстановления звука.
-