

МСЭ-R

Сектор радиосвязи МСЭ

Рекомендация МСЭ-R BS.1770-3
(08/2012)

Алгоритмы измерения громкости звуковых программ и истинного пикового уровня звукового 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 Приложении 1 к Резолюции МСЭ-R 1. Формы, которые владельцам патентов следует использовать для представления патентных заявлений и деклараций о лицензировании, представлены по адресу: <http://www.itu.int/ITU-R/go/patents/en>, где также содержатся Руководящие принципы по выполнению общей патентной политики МСЭ-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	Подвижная спутниковая служба, спутниковая служба радиоопределения, любительская спутниковая 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, 2012 г.

© ITU 2012

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.1770-3*

**Алгоритмы измерения громкости звуковых программ
и истинного пикового уровня звукового сигнала**

(Вопрос МСЭ-R 2/6)

(2006-2007-2011-2012)

Сфера применения

В настоящей Рекомендации описываются алгоритмы измерения звуковых сигналов для целей определения субъективной громкости программ и истинного пикового уровня сигналов.

Ассамблея радиосвязи МСЭ,

учитывая,

- a) что современные методы цифровой передачи звука предоставляют чрезвычайно обширный динамический диапазон;
- b) что использование современных цифровых методов производства и передачи звука обуславливает смешение монофонических, стереофонических и многоканальных форматов и что звуковые программы производятся во всех этих форматах;
- c) что слушатели желали бы, чтобы субъективная громкость звуковых программ была единой для различных источников и типов программ;
- d) что имеется множество методов измерения уровней громкости, но существующие методы измерений, используемые при производстве программ, не обеспечивают индикации субъективной громкости;
- e) что для целей регулировки громкости при обмене программами, с тем чтобы снизить уровень недовольства аудитории, важно использовать единый рекомендованный алгоритм объективной оценки субъективной громкости;
- f) что будущие комплексные алгоритмы, базирующиеся на психоакустических моделях, могут обеспечить более точные объективные измерения громкости для широкого спектра звуковых программ;
- g) что перегрузка цифрового носителя происходит скачкообразно, и поэтому следует избегать даже кратковременной его перегрузки,

учитывая далее,

- h) что пиковые уровни сигналов могут увеличиваться вследствие широко применяемых процессов, таких как фильтрация или уменьшение скорости передачи в битах;
- j) что существующие методы измерений не отражают истинного пикового уровня, достигаемого цифровым сигналом, поскольку истинное пиковое значение может появляться в интервале между отсчетами;
- k) что возможности обработки цифрового сигнала позволяют реализовать алгоритм, обеспечивающий оценку истинного пикового уровня сигнала с высокой точностью;
- l) что использование алгоритма индикации истинного пика делает возможным точную индикацию запаса по уровню между пиковым уровнем цифрового звукового сигнала и уровнем срезания пиков,

* В мае 2011 года 6-я Исследовательская комиссия по радиосвязи внесла редакционные поправки в настоящую Рекомендацию.

рекомендует

1 применять, в случае необходимости объективного измерения громкости звукового канала или программы в целях содействия доставке программ и обмена ими, алгоритм, описание которого содержится в Приложении 1;

2 обеспечивать возможность использования алгоритма, описание которого содержится в Приложении 1, в качестве основы для методов индикации громкости программы при производстве и окончательном монтаже программ;

3 применять, в случае необходимости индикации истинного пикового уровня цифрового звукового сигнала, метод измерений, в основу которого положены руководящие принципы, изложенные в Приложении 2, или метод, обеспечивающий аналогичные или лучшие результаты,

ПРИМЕЧАНИЕ 1. – Пользователей следует предупреждать, что измеренная громкость является оценкой субъективной громкости и связана с некоторой неопределенностью, обусловливаемой свойствами слушателей, звукового материала и условиями прослушивания.

далее рекомендует

1 учитывать возможную необходимость обновления настоящей Рекомендации, в случае если новые алгоритмы измерения громкости будут существенно превосходить по качеству работы алгоритм, описание которого содержится в Приложении 1.

ПРИМЕЧАНИЕ 2. – В целях обеспечения соответствия измерителей, используемых при проверке в соответствии с настоящей Рекомендацией, может применяться тестовый материал из набора, описанного в Отчете МСЭ-R BS.2217.

Приложение 1

Описание алгоритма объективного многоканального измерения громкости

В настоящем Приложении описан алгоритм многоканального измерения громкости.

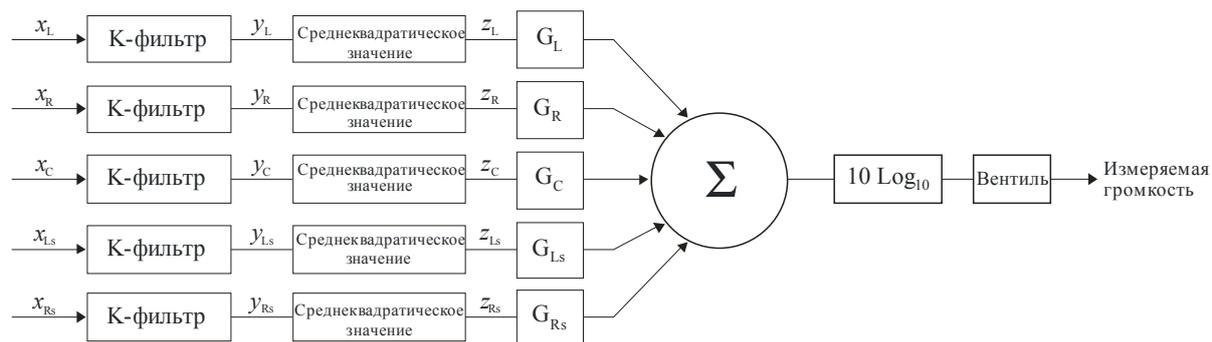
Алгоритм состоит из четырех этапов:

- К-взвешивание частоты;
- расчет среднеквадратического значения для каждого канала;
- суммирование с взвешиванием по каналам (тыловые каналы имеют больший вес, а каналы LFE исключаются);
- стробирование по блокам 400 мс (с перекрытием на 75%) с использованием двух пороговых значений:
 - первое на -70 LKFS;
 - второе на -10 дБ относительно уровня, измеренного после применения первого порогового значения.

На рисунке 1 представлена блок-схема различных компонентов алгоритма. Для удобства описания алгоритма на различных точках тракта потока сигнала предусмотрены метки. На блок-схеме показаны входы пяти основных каналов (левого, центрального, правого, левого тылового и правого тылового); это позволяет осуществлять контроль программ, имеющих от одного до пяти каналов. Если программа имеет менее пяти каналов, некоторые входы не используются. Канал для низкочастотных звуковых эффектов (LFE) данным измерением не охвачен.

РИСУНОК 1

Упрощенная блок-схема алгоритма многоканального измерения громкости

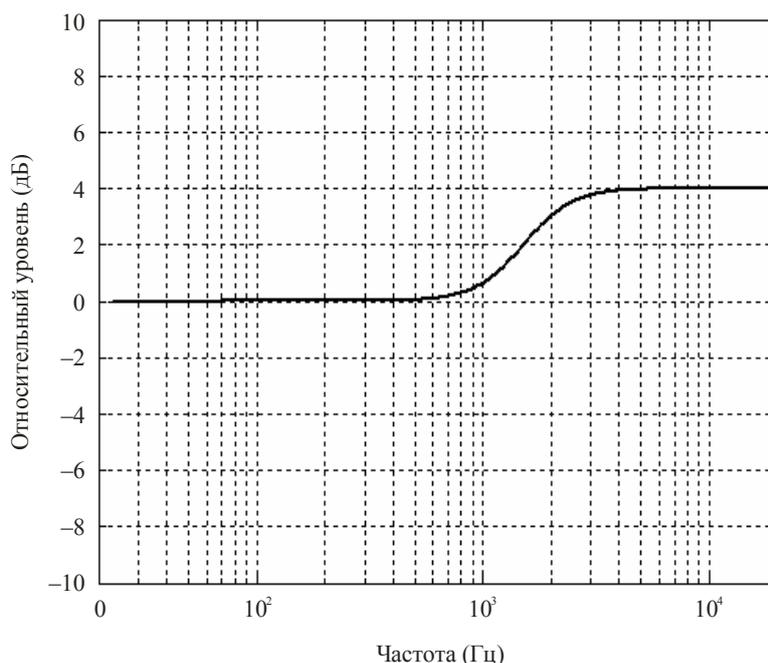


BS.1770-01

На первом этапе работы алгоритма применяется 2-ступенчатая предварительная фильтрация¹ сигнала. Первая ступень предварительной фильтрации обеспечивает учет акустического воздействия головы, при этом голова моделируется как жесткая сфера. Частотная характеристика показана на рисунке 2.

РИСУНОК 2

Частотная характеристика первой ступени предварительного фильтра, используемой для обеспечения учета акустического воздействия головы



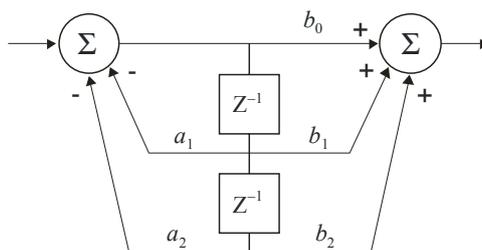
BS.1770-02

Первая ступень предварительного фильтра определяется как фильтр, показанный на рисунке 3, коэффициенты которого указаны в таблице 1.

¹ Фильтр К-взвешивания осуществляет две ступени фильтрации; первая ступень – сглаживающий фильтр, вторая ступень – фильтр верхних частот.

РИСУНОК 3

Диаграмма потока сигналов как фильтр 2-го порядка



BS.1770-03

ТАБЛИЦА 1

**Коэффициенты фильтра для первой ступени предварительного фильтра
в целях моделирования сферической головы**

		b_0	1,53512485958697
a_1	-1,69065929318241	b_1	-2,69169618940638
a_2	0,73248077421585	b_2	1,19839281085285

Эти коэффициенты фильтра соответствуют частоте дискретизации 48 кГц. Для реализаций с другой частотой дискретизации потребуются другие значения коэффициентов, которые должны выбираться так, чтобы обеспечивалась частотная характеристика, аналогичная характеристике указанного фильтра при частоте дискретизации 48 кГц. Присущая имеющемуся аппаратному обеспечению точность может обусловить необходимость квантования значений этих коэффициентов. Испытания показали, что алгоритм нечувствителен к незначительным изменениям этих коэффициентов.

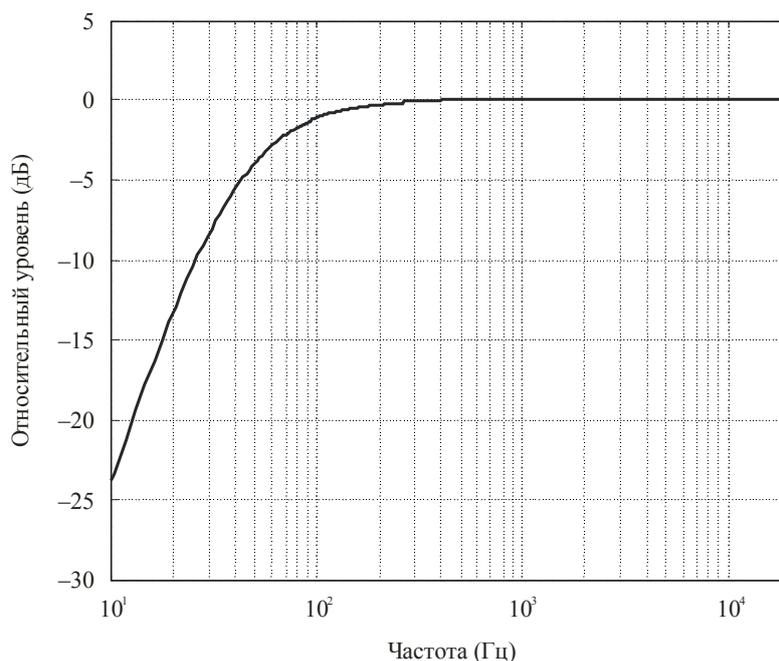
На второй ступени фильтра применяется простой фильтр верхних частот, как показано на рисунке 4.

Взвешивающая кривая этой ступени определяется как фильтр 2-го порядка, показанный на рисунке 3, коэффициенты которого указаны в таблице 2.

Эти коэффициенты фильтра соответствуют частоте дискретизации 48 кГц. Для реализаций с другой частотой дискретизации потребуются другие значения коэффициентов, которые должны выбираться так, чтобы обеспечивалась частотная характеристика, аналогичная характеристике указанного фильтра при частоте дискретизации 48 кГц.

РИСУНОК 4

Взвешивающая кривая второй степени



BS.1770-04

ТАБЛИЦА 2

Коэффициенты фильтра для взвешивающей кривой второй степени

		b_0	1,0
a_1	-1,99004745483398	b_1	-2,0
a_2	0,99007225036621	b_2	1,0

Энергия, среднеквадратическое значение отфильтрованного входного сигнала на интервале измерений T , определяется как:

$$z_i = \frac{1}{T} \int_0^T y_i^2 dt, \quad (1)$$

где y_i – входной сигнал (прошедший через две степени предварительного фильтра, описанные выше), а $i \in I$, где $I = \{L, R, C, Ls, Rs\}$ – совокупность входных каналов.

Громкость в интервале измерений T определяется как:

$$\text{Громкость, } L_K = -0,691 + 10 \log_{10} \sum_i G_i \cdot z_i \quad \text{LKFS}, \quad (2)$$

где G_i – коэффициенты взвешивания для отдельных каналов.

Для расчета измеренного значения стробированной громкости интервал T делится на множество перекрывающихся интервалов стробирующих блоков. Стробирующий блок – это множество непрерывных отсчетов звукового сигнала длительностью $T_g = 400$ мс относительно ближайшего отсчета. Перекрытие каждого стробирующего блока должно составлять 75% длительности стробирующего блока.

Интервал измерений должен быть ограничен так, чтобы он оканчивался в конце стробирующего блока. Стробирующие блоки, не завершённые на момент конца интервала измерений, не используются.

Энергия, среднеквадратическое значение j -го стробирующего блока i -го входного канала в интервале T , составляет:

$$z_{ij} = \frac{1}{T_g} \int_{T_g \cdot j \cdot шаг}^{T_g \cdot (j \cdot шаг + 1)} y_i^2 dt \quad \text{где} \quad шаг = 1\text{-перекрывание}$$

и $j \in \left\{ 0, 1, 2, \dots, \frac{T - T_g}{T_g \cdot шаг} \right\}$.

(3)

Громкость j -го стробирующего блока определяется как:

$$l_j = -0,691 + 10 \log_{10} \sum_i G_i \cdot z_{ij}. \quad (4)$$

Для порогового значения стробирования Γ существует множество индексов стробирующего блока $J_g = \{j : l_j > \Gamma\}$, где громкость стробирующего блока выше порогового уровня стробирования. Число элементов в J_g равно $|J_g|$.

Стробируемая громкость в интервале измерения T определяется таким образом как:

$$\text{Стробируемая громкость, } L_{KG} = -0,691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) LKFS. \quad (5)$$

Двухступенчатый процесс используется для выполнения стробируемых измерений, сначала с абсолютным пороговым значением, затем с относительным пороговым значением. Относительное пороговое значение Γ_r рассчитывается путем измерения громкости с использованием абсолютного порогового значения, $\Gamma_a = -70$ LKFS и уменьшения результата на 10, таким образом:

$$\Gamma_r = -0,691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) - 10 \text{ LKFS},$$

где:

$$\begin{aligned} J_g &= \{j : l_j > \Gamma_a\} \\ \Gamma_a &= -70 \text{ LKFS} \end{aligned} \quad (6)$$

Стробируемая громкость далее может быть рассчитана с использованием Γ_r :

$$\text{Стробируемая громкость, } L_{KG} = -0,691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) LKFS,$$

где:

$$J_g = \{j : l_j > \Gamma_r\}. \quad (7)$$

Взвешивание частоты в этом показателе, который генерируется предварительным фильтром (соединение фильтра первой ступени для компенсации акустических эффектов головы и фильтра второй ступени – взвешивание RLB) называется "К-взвешиванием". За полученным цифровым значением громкости, рассчитанным в уравнении (2), следует указывать обозначение "LKFS". Это обозначение расшифровывается как "громкость, взвешенная по К, относительно номинальной полной шкалы". Единица LKFS эквивалентна децибелу, поскольку увеличение уровня сигнала на 1 дБ приведет к увеличению показателя громкости на 1 LKFS.

Если ко входу левого, центрального или правого каналов будет приложена синусоидальная волна, уровень которой составляет 0 дБ полной шкалы с частотой 1 кГц, указанная громкость будет равняться –3,01 LKFS.

Весовые коэффициенты для каждого канала приведены в таблице 3.

ТАБЛИЦА 3

Весовые значения для отдельных звуковых каналов

Канал	Весовое значение, G_i
Левый (G_L)	1,0 (0 дБ)
Правый (G_R)	1,0 (0 дБ)
Центральный (G_C)	1,0 (0 дБ)
Левый тыловой (G_{Ls})	1,41 (~ +1,5 дБ)
Правый тыловой (G_{Rs})	1,41 (~ +1,5 дБ)

Следует отметить, что, хотя данный алгоритм доказал свою эффективность в отношении звуковых программ, которые являются типовым радиовещательным контентом, в целом он не пригоден для оценки субъективной громкости чистых тонов.

**Дополнение 1
к Приложению 1
(Информативное)**

Описание и разработка алгоритма многоканального измерения

В настоящем Дополнении описывается недавно разработанный алгоритм, предназначенный для объективного измерения воспринимаемой громкости звуковых сигналов. Этот алгоритм может использоваться для точного измерения громкости моно-, стерео- и многоканальных сигналов. Основным преимуществом предлагаемого алгоритма является его простота, делающая реализацию алгоритма весьма низкочастотной. В Дополнении также представлены результаты официальных субъективных испытаний, проведенных для создания базы субъективных данных, которая использовалась для оценки качества работы алгоритма.

1 Введение

Существует множество приложений, для которых требуются измерение и контроль воспринимаемой громкости звукового сигнала. Примерами этого являются телевизионные и радиовещательные приложения, характер и контент звукового материала которых быстро меняются. В таких приложениях звуковой контент может представлять собой постоянно и быстро меняющуюся музыку, речь и звуковые эффекты или их комбинацию. Такие изменения в контенте программы могут приводить к значительным изменениям субъективной громкости. Кроме того, к сигналам, которые могут оказывать значительное воздействие на воспринимаемую громкость сигнала, часто применяются различные формы динамической обработки. Разумеется, проблема субъективной

громкости имеет существенное значение для индустрии звукозаписи, где широко используется динамическая обработка с целью максимально возможного увеличения воспринимаемой громкости записи.

На протяжении ряда последних лет Рабочая группа 6Р Сектора радиосвязи ведет работу по определению методов объективного измерения воспринимаемой громкости программного материала, типового для радиовещательных приложений. Первый этап работы МСЭ-R, в ходе которого были рассмотрены алгоритмы объективного измерения громкости монофонического звука, а также взвешенное среднеквадратическое измерение, *Leq*(RLB), был посвящен обеспечению наилучшего качества воспроизведения монофонических сигналов [Soulodre, 2004].

Совершенно очевидно, что для радиовещательных приложений требуется измеритель громкости, который может работать с моно-, стерео- и многоканальными сигналами. В настоящем документе предлагается новый алгоритм измерения громкости, который успешно функционирует в случае моно-, стерео- и многоканальных звуковых сигналов. Основой предлагаемого алгоритма является прямое расширение алгоритма *Leq*(RLB). Кроме того, новый многоканальный алгоритм по-прежнему характеризуется простотой вычислений, свойственной монофоническому алгоритму *Leq*(RLB).

2 Базовая информация

На первом этапе исследований МСЭ-R был разработан метод субъективных испытаний для анализа восприятия громкости материала типовой монофонической программы [Soulodre, 2004]. Субъективные испытания были проведены в пяти точках земного шара, и их целью было создание базы субъективных данных для оценки качества работы возможных алгоритмов измерения громкости. В ходе этих испытаний участники сопоставляли громкость различных монофонических звуковых записей с громкостью эталонной записи. Звуковые записи были взяты из реального радиовещательного материала (телевидения и радио).

Совместно с этими испытаниями семь инициаторов представили в Лабораторию слухового восприятия Исследовательского центра связи (Канада) десять внедренных измерителей/алгоритмов измерения монофонической громкости для их оценки.

Кроме того, в работах Сулодра (Soulodre) были представлены еще два базовых алгоритма измерения громкости в качестве основы в отношении показателей работы алгоритма [Soulodre, 2004]. Эти два метода объективных измерений включали простую функцию спектрального взвешивания, за которой следовал блок измерения среднеквадратического значения. Один из двух методов измерений, (*Leq*(RLB)), использует взвешивающую кривую верхних частот, называемую В-кривой с коррекцией по нижним частотам (RLB).

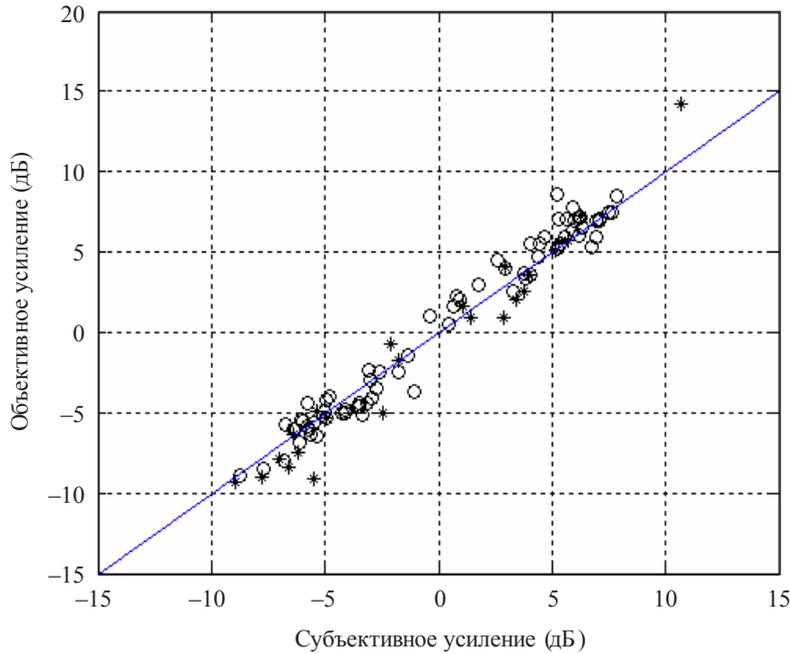
Другой метод измерений, *Leq*, заключается в измерении невзвешенного среднеквадратического значения.

На рисунке 5 представлены результаты начального исследования МСЭ-R измерителя громкости *Leq*(RLB). По горизонтальной оси показана относительная субъективная громкость, выведенная из базы субъективных данных, по вертикальной оси – громкость, прогнозируемая при измерении *Leq*(RLB). Каждая точка на графе представляет результат по одной из использованных в испытаниях звуковых тестовых последовательностей. Незаштрихованные кружки относятся к звуковым последовательностям на базе речи, а звездочки – к последовательностям, не содержащим речи. Как можно видеть, точки данных плотно группируются вокруг диагонали, отражая очень хорошее качество работы измерителя *Leq*(RLB).

Измеритель *Leq*(RLB) был признан имеющим наилучшие показатели качества по сравнению со всеми исследовавшимися измерителями (хотя в пределах статистической значимости некоторые измерители на базе психоакустики тоже показывали хорошее качество). Оказалось, что измеритель *Leq* работал почти так же хорошо, как RLB. Полученные результаты означают, что для типового монофонического радиовещательного материала простой метод измерения громкости на основе энергии является столь же устойчивым, как и более сложные методы измерений, которые могут включать детализированные модели восприятия.

РИСУНОК 5

Результаты измерителя монофонической громкости $Leq(RLB)$ в сравнении с результатами субъективных измерений ($r = 0,982$)



BS.1770-05

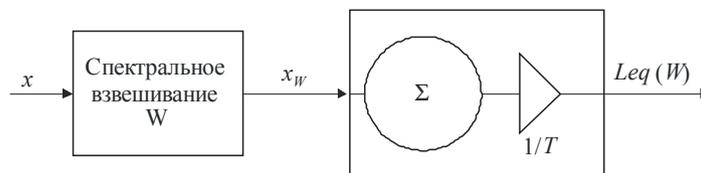
3 Схема алгоритма $Leq(RLB)$

Алгоритм измерения громкости $Leq(RLB)$ специально разрабатывался так, чтобы быть очень простым. Блок-схема алгоритма $Leq(RLB)$ показана на рисунке 6. Она содержит фильтр верхних частот, за которым следует устройство усреднения энергии по времени. Сигнал с выхода фильтра поступает в блок обработки, который суммирует энергию и вычисляет среднюю величину по времени.

Задачей фильтра является обеспечение определенного взвешивания, соответствующего чувственному восприятию, спектрального содержимого сигнала. Одним из преимуществ использования этой базовой структуры для измерений громкости является возможность выполнения всей обработки простыми блоками временных интервалов, в отношении которых требуются весьма низкие объемы вычислений.

РИСУНОК 6

Блок-схема простого метода измерений громкости на основе энергии



BS.1770-06

Алгоритм $Leq(RLB)$, показанный на рисунке 6, – это просто вариант со спектральным взвешиванием метода измерения *Эквивалентного уровня звукового давления* (Leq). Leq определяется следующим образом:

$$Leq(W) = 10 \log_{10} \left[\frac{1}{T} \int_0^T \frac{x_W^2}{x_{Ref}^2} dt \right] \quad \text{дБ}, \quad (3)$$

где:

- x_W : сигнал на выходе взвешивающего фильтра;
- x_{Ref} : некоторый эталонный уровень;
- T : длина звуковой последовательности.

Символ W в $Leq(W)$ означает спектральное взвешивание, которым в данном случае являлась В-кривая с коррекцией по нижним частотам (RLB).

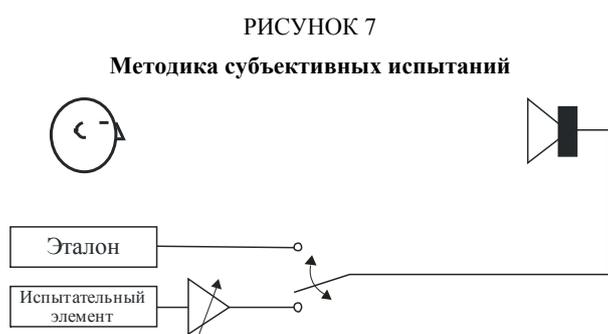
4 Субъективные испытания

Для оценки возможных методов многоканальных измерений громкости необходимо было провести официальные субъективные испытания, с тем чтобы сформировать базу субъективных данных. Затем потенциальные алгоритмы измерения громкости могли быть оценены по своей способности прогнозировать результаты субъективных испытаний. База данных обеспечивала значения воспринимаемой громкости для обширного материала моно-, стерео- и многоканальных программ. Программный материал, который использовался в ходе испытаний, был взят из реальных радио- и телевизионных передач по всему миру, а также из записей на CD и DVD. Звуковые последовательности включали музыку, телевизионные и кинопрограммы, материалы спортивных мероприятий, программы новостей, звуковые эффекты и рекламу. В эти последовательности были включены сегменты речи на различных языках.

4.1 Схема субъективных испытаний

Субъективные испытания заключались в сопоставлении уровней громкости. Участники прослушивали широкий диапазон типового программного материала и регулировали уровень громкости каждого тестового элемента до совпадения его воспринимаемой громкости с громкостью эталонного сигнала (см. рисунок 7).

Эталонный сигнал воспроизводился всегда на уровне 60 дБА, это значение было определено Бенджамином как типовой уровень прослушивания при просмотре телевизионных программ в реальных домашних условиях [Benjamin, 2004].

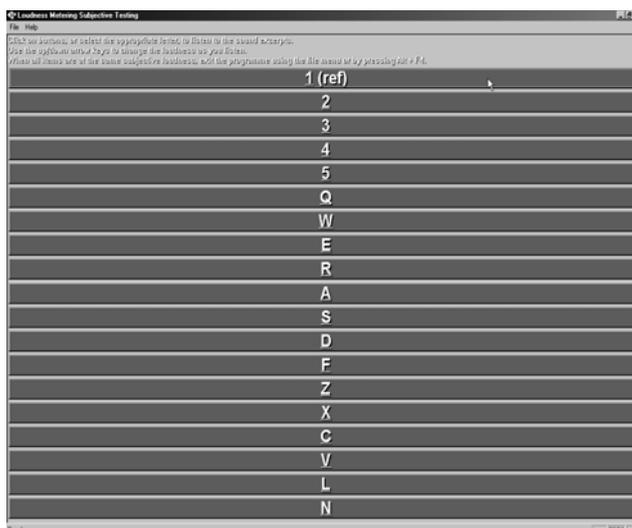


BS.1770-07

Программно реализованная система для проведения субъективных испытаний многоканальных измерений, разработанная и предоставленная Австралийской радиовещательной корпорацией, позволяла слушателю моментально переключаться с одного испытательного элемента на другой и регулировать уровень (громкость) каждого элемента. Снимок экрана программного обеспечения для проведения испытаний показан на рисунке 8. Уровень испытательных элементов можно было регулировать с шагом 0,25 дБ. Выбор кнопки, имеющей метку "1", обеспечивал доступ к эталонному сигналу. Уровень эталонного сигнала сохранялся постоянным.

РИСУНОК 8

Интерфейс пользователя системы для проведения субъективных испытаний



BS.1770-08

С помощью компьютерной клавиатуры участник испытаний выбирал данный испытательный элемент и регулировал его уровень до совпадения громкости этого элемента с громкостью эталонного сигнала. Участники могли моментально переключаться на любой испытательный элемент, выбирая для этого соответствующую клавишу. В течение испытаний последовательности воспроизводились непрерывно (замкнутая петля). Программное обеспечение фиксировало установки усиления для каждого испытательного элемента, определенные участником. Таким образом, в результате субъективных испытаний формировался набор значений усиления (в децибелах), которые требовались для совпадения громкости каждой тестовой последовательности с громкостью эталонной последовательности. Это позволяло напрямую определять относительную громкость каждого испытательного элемента.

До проведения официальных испытаний вслепую каждый участник посетил учебное занятие, на котором знакомился с программным обеспечением испытаний и с задачей, которую он должен был выполнять в ходе эксперимента. Поскольку многие испытательные элементы содержали смесь речи и других звуков (музыка, фоновые шумы и т. д.), участникам специально указывалось на необходимость сопоставления громкости общего сигнала, а не речевого компонента сигналов.

В ходе официальных испытаний вслепую порядок представления последовательностей каждому участнику был произвольным. Таким образом, всем участникам последовательности представлялись в разном порядке. Это было сделано во избежание любой возможной систематической ошибки, обусловленной влиянием порядка.

4.2 База субъективных данных

База субъективных данных, которая использовалась для оценки качества предлагаемого алгоритма, состояла фактически из трех различных баз данных. Наборы данных были созданы по результатам трех независимых субъективных испытаний, проведенные в течение нескольких лет.

Первый набор данных составляли результаты первоначального исследования МСЭ-R, в ходе которого участники сопоставляли воспринимаемую громкость 96 монофонических звуковых последовательностей. Для создания этого набора данных субъективные испытания проводились в пяти различных точках земного шара и в них участвовали 97 слушателей. Испытательные последовательности и эталонный элемент выбирала группа из трех человек, в которую вошли члены SRG3 PG 6P (Специальная группа Докладчика). Эталонная последовательность в этом испытании содержала речь на английском языке, произносимую женским голосом. Воспроизведение последовательностей осуществлялось через один громкоговоритель, размещенный непосредственно напротив слушателя.

После первоначального монофонического исследования МСЭ-R ряд инициаторов, представивших алгоритмы, выдвинули предположение, что диапазон и тип сигналов, которые использовались в субъективных испытаниях, не были достаточно широкими. Они далее предположили, что в этом заключалась причина того, что простой алгоритм *Leq*(RLB) на основе энергии превзошел по качеству работы все другие алгоритмы.

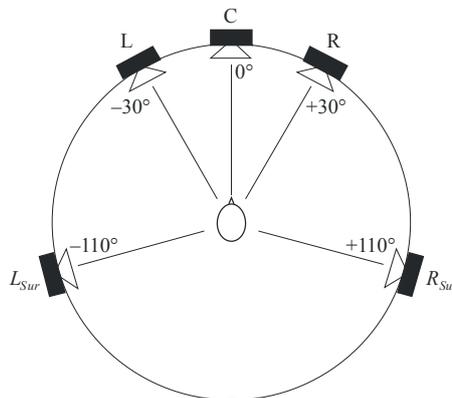
Для того чтобы рассеять эти сомнения, поднявшим вопрос было предложено представить новую звуковую последовательность для следующего раунда субъективных испытаний. Им было рекомендовано представить монофонические последовательности, которые, по их мнению, явились бы наиболее сложными для алгоритма *Leq*(RLB). Новые последовательности представили только двое из высказавших сомнения.

Официальные субъективные испытания с этими новыми последовательностями были проведены в Лаборатории слухового восприятия Исследовательского центра связи (Канада). Громкость 96 монофонических последовательностей определяли 20 участников испытаний. В этих испытаниях использовалась та же методика субъективных испытаний, что и при создании первого набора данных, и тот же эталонный сигнал. Результаты этих испытаний составили второй набор данных базы субъективных данных.

Третий набор данных состоял из значений громкости для 144 звуковых последовательностей. Испытательные последовательности содержали 48 монофонических элементов, 48 стереофонических элементов и 48 многоканальных элементов. Кроме того, половина монофонических элементов воспроизводилась через центральный канал (моно), а другая половина монофонических элементов – через левый и правый громкоговорители (двойное моно). Это было сделано с тем, чтобы учитывать два присущих человеку различных способа прослушивания монофонического сигнала. Для этого испытания эталонный сигнал содержал речь на английском языке, произносимую женским голосом, со стереофоническим акустическим окружением и музыкальным фоном низкого уровня. В испытаниях участвовали 20 человек и использовалась схема расположения громкоговорителей, определенная в Рекомендации МСЭ-R BS.775 и представленная на рисунке 9.

РИСУНОК 9

Схема расположения громкоговорителей, которая использовалась для формирования третьего набора данных



BS.1770-09

Два первых набора данных были ограничены монофоническими испытательными последовательностями, поэтому создание звукового образа не являлось влияющим фактором. При создании третьего набора данных, который включал также стерео- и многоканальные последовательности, создание образа являлось важным соображением, которое следовало учитывать. Представлялось, что создание образа и акустическое окружение в последовательности по всей вероятности могут оказывать значительное воздействие на воспринимаемую громкость этой последовательности. Таким образом, для включения широкого диапазона видов создания образа (например, центральное панорамирование – резкая граница звука между левым/правым каналами, источники звука напротив – источники звука вокруг) и меняющегося объема акустического окружения (например, сухое звучание – реверберация) были выбраны стерео- и многоканальные последовательности.

Тот факт, что участники испытаний должны были одновременно сопоставлять громкость сигналов моно-, двойного моно-, стерео- и многоканальных сигналов, означал, что эти испытания являлись по своему существу значительно сложнее испытаний по составлению предыдущих наборов данных, которые были ограничены монофоническими сигналами. Эта сложность усугублялась использованием различных видов создания образа и меняющегося объема акустического окружения. Существовали определенные сомнения, что вследствие всех этих факторов задача для участников испытаний окажется непосильной. К счастью, предварительные испытания показали, что задача является выполнимой, и все 20 участников оказались в состоянии обеспечивать достоверные результаты.

5 Схема алгоритма многоканального измерения громкости

Как было указано выше, алгоритм $Leq(RLB)$ был создан для работы с монофоническими сигналами, и ранее проведенное исследование показало высокую эффективность алгоритма для этой задачи. Разработка алгоритма многоканального измерения громкости обуславливает ряд дополнительных сложностей. Ключевым требованием, предъявляемым к многоканальному алгоритму, является его способность работать также эффективно с сигналами моно-, двойного моно- и стереосигналами. Это значит, что эти форматы должны рассматриваться как особые случаи многоканального сигнала (хотя они и являются весьма общими случаями).

В настоящем исследовании предполагалось, что многоканальные сигналы соответствуют стандартной конфигурации каналов 5.1 Рекомендации МСЭ-R BS.775. Учет канала LFE не осуществлялся.

В многоканальном измерителе громкости громкость каждого отдельного звукового канала измеряется независимо с помощью монофонического алгоритма $Leq(RLB)$, как показано на рисунке 10. Вместе с тем в каждом канале до выполнения измерения $Leq(RLB)$ выполняется предварительная фильтрация.

РИСУНОК 10

Блок-схема предлагаемого многоканального измерителя громкости



BS.1770-10

Задачей предварительного фильтра является обеспечение учета акустического воздействия, которое оказывает на входной сигнал головы. При этом голова моделируется как жесткая сфера. В каждом канале используется один и тот же предварительный фильтр. Далее результирующие значения громкости взвешиваются (G_i) в соответствии с углами прихода сигнала, а затем суммируются (в линейной области) для получения составного значения измеряемой громкости. Взвешивание используется с тем, чтобы учесть тот факт, что звуки, поступающие к слушателю сзади, могут восприниматься им как более громкие по сравнению со звуками, поступающими от источника, расположенного напротив. Сочетание "предварительного фильтра" и "фильтра RLB" на рисунке 10 относится к K-взвешиванию, как указано в основной части Приложения 1, выше.

Ключевым преимуществом предлагаемого алгоритма многоканального измерения громкости является его простота. Алгоритм полностью составляют очень простые базовые блоки обработки сигнала, которые могут быть легко и быстро реализованы с помощью недорогого аппаратного обеспечения. Другим ключевым преимуществом алгоритма является его расширяемость. Учитывая,

что во всех каналах применяется идентичная обработка, очень просто реализовать измеритель, который может работать с любым количеством каналов от 1 до N . Кроме того, поскольку сигналы от каждого канала суммируются как значения громкости, а не как уровни сигнала, алгоритм не зависит от межканальной фазы или корреляции. Это делает предлагаемое измерение громкости значительно более универсальным и устойчивым.

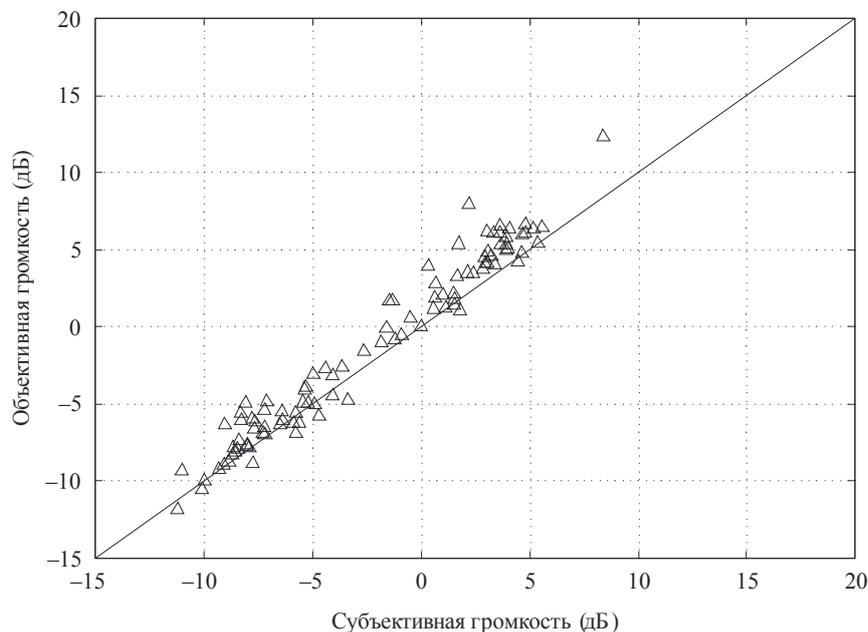
6 Оценка многоканального алгоритма

Использованные для трех наборов данных 336 звуковых последовательностей были обработаны с помощью предлагаемого многоканального алгоритма, и были зарегистрированы прогнозируемые значения громкости. По итогам этого процесса можно было оценить общий уровень эффективности работы данного алгоритма на основании совпадения прогнозируемых значений и реальных субъективных значений, полученных в ходе официальных субъективных испытаний.

На рисунках 11, 12 и 13 в графической форме показана эффективность работы предлагаемого измерителя громкости по трем наборам данных. На каждом рисунке по горизонтальной оси отмечены значения субъективной громкости каждой звуковой последовательности в наборе данных. По вертикальной оси показаны значения объективной громкости, прогнозируемые предлагаемым измерителем громкости. Каждая точка на графике представляет результат для отдельной звуковой последовательности. Следует заметить, что в случае идеального алгоритма объективных измерений все точки попали бы на диагональ, крутизна которой равна 1 и которая проходит через начало координат (как показано на рисунках).

РИСУНОК 11

Результаты для первого (моnofонического) набора данных ($r = 0,979$)



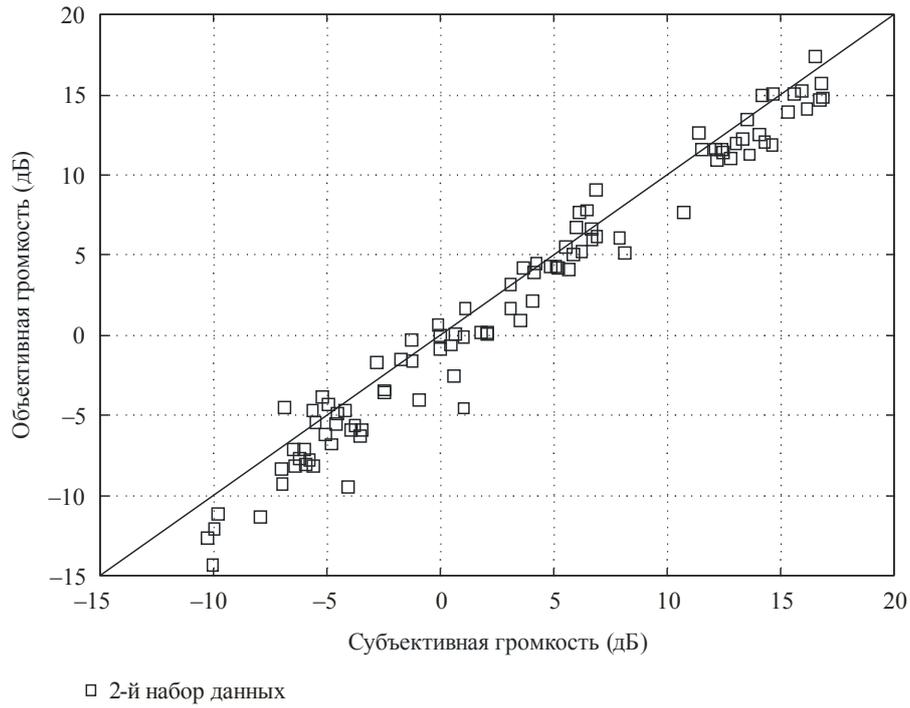
△ 1-й набор данных

BS.1770-11

Как видно на рисунке 11, предлагаемый алгоритм многоканального измерения громкости работает очень хорошо, прогнозируя результаты для первого (моnofонического) набора данных. Корреляция между значениями субъективной громкости и измерениями объективной громкости составляет $r = 0,979$.

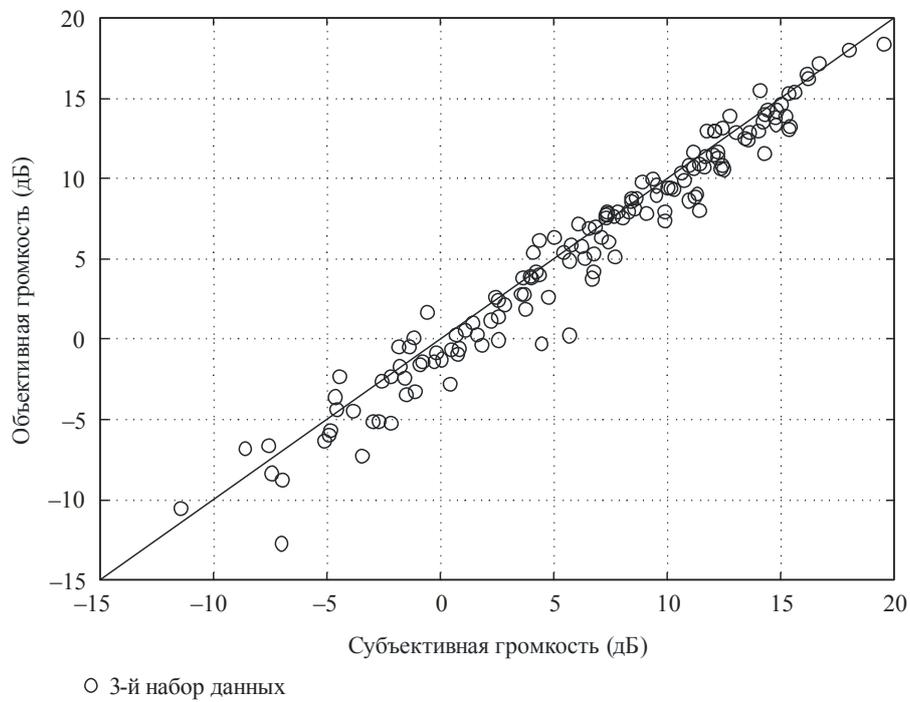
Как видно на рисунке 12, корреляция между значениями субъективной громкости и измерениями объективной громкости для второго набора данных также весьма удовлетворительная ($r = 0,985$). Интересно отметить, что около половины последовательностей в этом наборе данных были музыкальными.

РИСУНОК 12

Результаты для второго (монофонического) набора данных ($r = 0,985$)

BS.1770-12

РИСУНОК 13

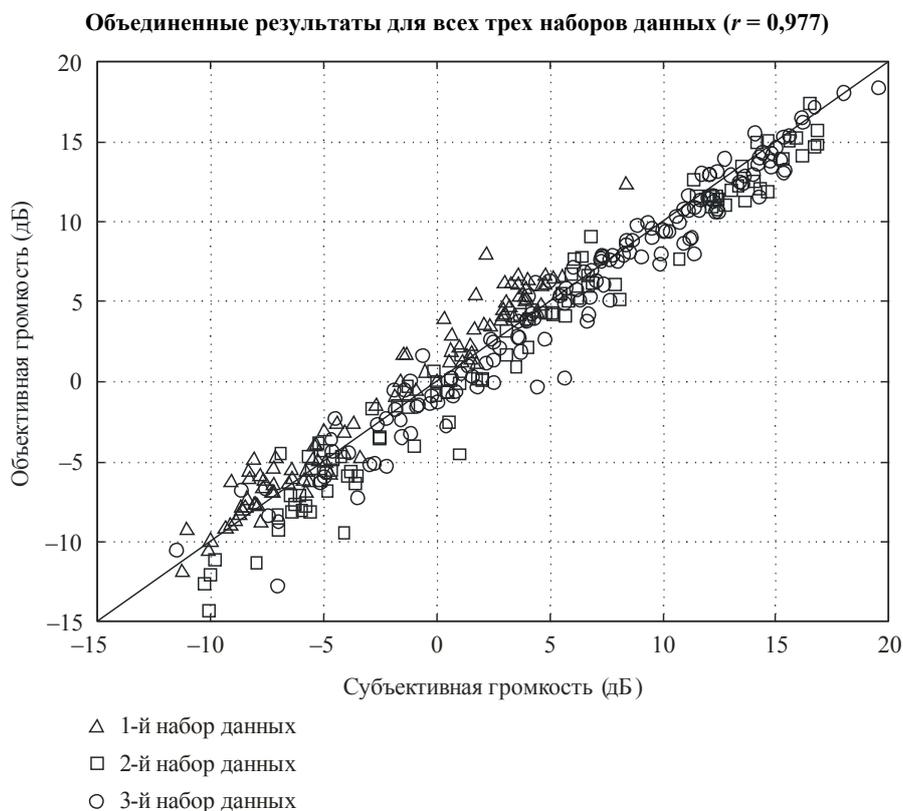
Результаты для третьего (моно-, стерео- и многоканального) набора данных ($r = 0,980$)

BS.1770-13

На рисунке 13 показаны результаты для третьего набора данных, который включал моно-, стерео- и многоканальные сигналы. Результаты для многоканальных сигналов, показанные на рисунках 13 и 14, относятся к указанному алгоритму, но взвешивание по тыловым каналам было установлено равным 4 дБ (исходное предложение) вместо 1,5 дБ (окончательная спецификация). Было проверено, что изменение значения с 4,0 дБ до 1,5 дБ не оказывает какого-либо существенного влияния на результаты. И опять алгоритм показал весьма удовлетворительное качество работы с корреляцией $r = 0,980$.

Полезно проверить эффективность работы алгоритма для всех 336 звуковых последовательностей, которые составили базу субъективных данных. Поэтому на рисунке 14 сведены вместе результаты по всем трем наборам данных. Можно видеть, что эффективность работы весьма удовлетворительная для всей базы субъективных данных, общая корреляция составляет $r = 0,977$.

РИСУНОК 14



BS.1770-14

Результаты этой оценки показывают, что алгоритм многоканального измерения громкости, основанный на измерении громкости $Leq(RLB)$, функционирует очень хорошо со всеми 336 последовательностями базы субъективных данных. База субъективных данных содержит широкий диапазон программного материала, включая музыку, телевизионные и кинопрограммы, материалы о спортивных мероприятиях, программы новостей, звуковые эффекты и рекламу. Также в эти последовательности были включены сегменты речи на различных языках. Кроме того, результаты показывают, что предлагаемый измеритель громкости хорошо работает в случае монофонических, двойных монофонических, стереофонических, а также многоканальных сигналов.

Справочные документы

BENJAMIN, E. [October, 2004] Preferred Listening Levels and Acceptance Windows for Dialog Reproduction in the Domestic Environment, 117th Convention of the Audio Engineering Society, San Francisco, Preprint 6233.

SOULODRE, G.A. [May 2004] Evaluation of Objective Loudness Meters, 116th Convention of the Audio Engineering Society, Berlin, Preprint 6161.

Приложение 2

Руководящие принципы для точного измерения "истинного пикового" уровня

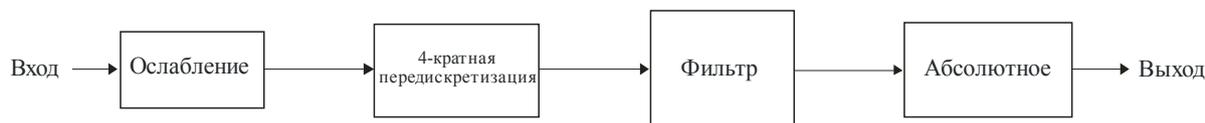
В настоящем Приложении представлен алгоритм оценки истинного пикового уровня одноканального цифрового звукового сигнала с линейной ИКМ. Частота дискретизации принята равной 48 кГц. Истинный пиковый уровень – это максимальное (положительное или отрицательное) значение сигнала в течение непрерывного интервала времени; это значение может быть выше наибольшего значения отсчета в рамках временной выборки при частоте дискретизации 48 кГц.

1 Резюме

Этапы обработки:

- 1 ослабление: ослабление на 12,04 дБ;
- 2 4-кратная передискретизация;
- 3 фильтр нижних частот;
- 4 абсолютное: абсолютное значение.

2 Блок-схема



BS.1770-15

3 Полное описание

Первый шаг заключается в применении ослабления на 12,04 дБ (сдвиг на 2 разряда). Цель этого шага – обеспечение запаса по уровню для последующей обработки сигнала, в которой применяется целочисленная арифметика. Этот шаг не требуется, если выполняются расчеты с плавающей запятой.

Фильтр с 4-кратной передискретизацией увеличивает частоту дискретизации сигнала с 48 кГц до 192 кГц. Этот вариант сигнала с повышенной частотой дискретизации более точно показывает реальную форму колебаний, представленную аудиовыборками. Предпочтительно использовать повышенные значения частоты дискретизации и коэффициенты передискретизации (см. Дополнение 1 к настоящему Приложению). Входные сигналы, имеющие повышенную частоту дискретизации, требуют пропорционально меньшей передискретизации (например, для входного сигнала с частотой дискретизации 96 кГц достаточно будет 2-кратной дискретизации).

Набор коэффициентов фильтра (для порядка 48, 4 этапа, с интерполяцией FIR), которые удовлетворяли бы этим требованиям, был бы следующим:

Этап 0	Этап 1	Этап 2	Этап 3
0,0017089843750	-0,0291748046875	-0,0189208984375	-0,0083007812500
0,0109863281250	0,0292968750000	0,0330810546875	0,0148925781250
-0,0196533203125	-0,0517578125000	-0,0582275390625	-0,0266113281250
0,0332031250000	0,0891113281250	0,1015625000000	0,0476074218750
-0,0594482421875	-0,1665039062500	-0,2003173828125	-0,1022949218750
0,1373291015625	0,4650878906250	0,7797851562500	0,9721679687500
0,9721679687500	0,7797851562500	0,4650878906250	0,1373291015625
-0,1022949218750	-0,2003173828125	-0,1665039062500	-0,0594482421875
0,0476074218750	0,1015625000000	0,0891113281250	0,0332031250000
-0,0266113281250	-0,0582275390625	-0,0517578125000	-0,0196533203125
0,0148925781250	0,0330810546875	0,0292968750000	0,0109863281250
-0,0083007812500	-0,0189208984375	-0,0291748046875	0,0017089843750

Абсолютное значение отсчетов получают путем инвертирования имеющих отрицательное значение отсчетов; в этой точке сигнал является однополярным, так как отрицательные значения заменены положительными значениями той же величины.

В измерениях, при которых следуют этим указаниям и в которых используется дискретизация с повышенной частотой не менее 192 кГц, следует указывать результат в единицах "дБ истинного пикового уровня". Это обозначение означает децибелы относительно полной шкалы в 100%, измерение истинного пикового уровня.

**Дополнение 1²
к Приложению 2
(Информативное)**

**Соображения относительно точного измерения
пика цифровых звуковых сигналов**

В чем заключается проблема?

Измерители пика в цифровых звуковых системах часто регистрируют "пиковый отсчет", а не "истинный пик".

Измерители пиковых отсчетов работают, как правило, путем сравнения абсолютного (среднего по модулю) значения каждого входного отсчета с текущим показанием измерителя; если значение нового отсчета выше, оно заменяет текущее показание; если нет, текущее показание умножается на постоянную, несколько меньшую единицы, для создания логарифмического затухания. Такие измерители широко распространены в силу простоты их реализации, но они не всегда регистрируют значение истинного пика звукового сигнала.

Следовательно, использование измерителя пикового отсчета в случаях, когда важно осуществлять точное измерение пиков программы, может привести к возникновению проблем. К сожалению, большинство цифровых измерителей пиков являются измерителями пиковых отсчетов, хотя это и не всегда очевидно для оператора.

Возникновение проблемы обусловлено тем, что реальные пиковые значения дискретизированного сигнала обычно появляются между отсчетами, а не точно в момент дискретизации, и как таковые правильно не регистрируются измерителем пиковых отсчетов.

Это является причиной нескольких известных аномалий измерителей пиковых отсчетов:

- *Неустойчивые значения пиков.* Часто отмечается, что при многократном воспроизведении аналоговой записи через цифровую систему с измерителем пиковых отсчетов значения программных пиков весьма сильно расходятся при каждом воспроизведении. Аналогично, если цифровая запись многократно воспроизводится через конвертор частоты дискретизации до измерения, регистрируемые пики точно так же различаются при каждом воспроизведении. Это происходит из-за того, что при каждом воспроизведении экземпляры отсчетов могут попадать в различные части истинного сигнала.
- *Непредвиденные перегрузки.* В силу того что дискретизированные сигналы могут содержать перегрузки даже в отсутствие отсчетов в пределах полной цифровой шкалы или близко к ней, индикация перегрузки измерителем пиковых отсчетов является ненадежной. Перегрузки могут вызвать срезание импульсов в последующих процессах, например в определенных цифро-аналоговых преобразователях или при преобразовании частоты дискретизации, даже если эти перегрузки ранее не были зарегистрированы измерителем пиковых отсчетов (и даже были неразличимы на слух при контроле их в этой точке).
- *Заниженное значение и биение измеряемых тонов.* Чистые тоны (такие как тоны настройки), близкие к целым множителям частоты дискретизации, могут давать заниженное значение или создавать постоянно меняющееся значение, даже при постоянной амплитуде тона.

Насколько серьезной может быть эта проблема?

В целом чем выше частота измеряемого дискретизированного пикового сигнала, тем больше потенциальная ошибка.

В случае непрерывных чистых тонов несложно продемонстрировать, например, заниженное на 3 дБ значение неудачно фазированного тона на четверти частоты дискретизации. Занижение значения тона на половине частоты дискретизации может быть фактически неограниченным;

² ПРИМЕЧАНИЕ 1. – Следующую ниже информацию для сведения представила Рабочая группа SC-02-01 по стандартам AES через Докладчика РГ 6J по радиосвязи по вопросам измерения громкости.

однако большинство цифровых звуковых сигналов не имеют на этой частоте достаточно энергии (в силу ее значительного исключения сглаживающими фильтрами в точке цифро-аналогового преобразования и поскольку в составе "реальных" звуков обычно не преобладают непрерывные высокие частоты).

Непрерывные тоны, не являющиеся близкими к низшим целым множителям частоты дискретизации, не дают заниженного значения в измерителях пиковых отсчетов, поскольку частота биения (разница между $n \cdot f_{tone}$ и f_s) весьма сравнима с величиной, обратной скорости затухания измерителя. Другими словами, момент дискретизации часто достаточно близок к истинному пику тона, чтобы измеритель не занижал его значение.

Вместе с тем для отдельных кратковременных звуков занижение значений этим механизмом не скрывается, поэтому чем более высокочастотным является содержимое кратковременного звука, тем больше потенциальное занижение значения. Для "реального" звука обычным является значительный объем высокочастотного содержимого в кратковременных звуках, и занижение их значения может обычно составлять несколько дБ.

В силу того что реальные звуки имеют спектр, совпадающий в направлении высоких частот, и это не меняется с увеличением частоты дискретизации, занижение значений измерителем пиковых отсчетов меньше на более высоких исходных частотах дискретизации.

Как разрешить эту проблему?

Для измерения значения истинного пика дискретизированного сигнала необходимо "передискретизировать" (или повысить частоту дискретизации) сигнала, фактически восстановив исходный сигнал между существующими отсчетами и увеличив таким образом частоту дискретизации сигнала. Это предложение вызывает сомнение: как мы можем восстановить информацию, которая, по-видимому, уже потеряна? На самом деле, теория дискретизации показывает, что можно это сделать, так как известно, что этот дискретизованный сигнал не содержит частот выше половины исходной частоты дискретизации.

Какой необходим коэффициент передискретизации? Чтобы узнать это, прежде необходимо ответить на два вопроса:

- Какова максимальная приемлемая ошибка вследствие занижения значения?
- Каково отношение наивысшей частоты, подлежащей измерению, к частоте дискретизации (максимальная "нормированная частота")?

Если эти критерии известны, становится возможным вычислить необходимый коэффициент передискретизации (даже не рассматривая еще детали реализации передискретизации) с помощью несложного метода "миллиметровой бумаги". Можно просто учитывать, какое занижение значения получится в результате пары отсчетов, симметрично появляющихся с передискретизованной скоростью по обеим сторонам пика синусоиды на нашей максимальной нормированной частоте. Это "худший случай" занижения значения.

Итак, для коэффициента передискретизации, n
 максимальной нормированной частоты, f_{norm}
 частоты дискретизации, f_s

можно видеть, что:

период дискретизации при передискретизованной скорости составляет $1/n \cdot f_s$

период максимальной нормированной частоты составляет $1/f_{norm} \cdot f_s$,

значит:

максимум занижения значения (дБ) составит $20 \cdot \log(\cos(2 \cdot \pi \cdot f_{norm} \cdot f_s / n \cdot f_s \cdot 2))$

(2 в знаменателе обусловлено вероятностью пропуска пика максимум на половину периода передискретизации)

или

максимум занижения значения (в дБ) = $20 \cdot \log(\cos(\pi \cdot f_{norm} / n))$.

Это уравнение использовалось для составления следующей ниже таблицы, которая, наиболее вероятно, охватывает диапазон, представляющий интерес:

Коэффициент передискретизации	Максимум занижения значения (дБ) $f_{norm} = 0,45$	Максимум занижения значения (дБ) $f_{norm} = 0,5$
4	0,554	0,688
8	0,136	0,169
10	0,087	0,108
12	0,060	0,075
14	0,044	0,055
16	0,034	0,042
32	0,008	0,010

Как должен быть реализован измеритель истинного пика?

Операция передискретизации выполняется путем включения отсчетов с нулевыми значениями между исходными отсчетами, с тем чтобы образовать поток данных с заданной передискретизованной скоростью и последующим применением "интерполирующего" фильтра нижних частот для исключения частот выше заданной максимальной f_{norm} . Если теперь применить алгоритм измерения пиковых отсчетов к передискретизированному сигналу, получим измеритель истинного пика с желаемым максимальным уровнем занижения значения.

Интересно рассмотреть реализацию такого передискретизатора. Принято выполнять такой фильтр нижних частот как симметричный фильтр с конечной импульсной характеристикой (FIR). При использовании таких фильтров для пропускания высококачественного звукового материала, например в (устаревших) цифро-аналоговых преобразователях передискретизации или преобразователях скорости дискретизации, необходимо вычислять большое число "отводов", с тем чтобы поддерживать очень малую неравномерность в полосе пропускания и добиваться наибольшего ослабления в полосе заграждения, а также узкой переходной полосы. Также должна поддерживаться большая длина слова для сохранения динамического диапазона и минимизации искажения.

Вместе с тем, поскольку мы не намерены прослушивать выход нашего передискретизатора, а только использовать его для отображения значений или построения гистограммы, вероятно, можно не предъявлять такие же требования к точности. До тех пор пока неравномерность в полосе пропускания, связанная с паразитными компонентами из заграждающей полосы, не ухудшает точности значений по сравнению с заданной нами, считаем ее удовлетворительной. Это в значительной степени сокращает количество необходимых отводов, хотя может сохраняться необходимость в обеспечении узкой переходной полосы, что зависит от заданной нами максимальной нормированной частоты. Аналогично можно ограничиться требованием, чтобы длина слова была достаточной для гарантии поддержания целевой точности вплоть до основания гистограммы, если только не требуется точный числовой результат для низких амплитуд.

Таким образом, возможна реализация соответствующего передискретизатора (может быть для большого числа каналов) на обычном недорогом процессоре DSP или FPGA, а может, и на более простом процессоре. С другой стороны, измерители с передискретизацией выполняются с использованием высокоточных чипов передискретизации, которые предназначены для использования в цифро-аналоговых преобразователях. Хотя это и в некоторой степени расточительное использование кремния и электроэнергии, устройства являются недорогими и легкодоступными.

Простейшим методом определения требуемого числа отводов и коэффициентов отводов, исходя из спецификации конкретного измерителя, является использование программы проектирования рекурсивных фильтров FIR, такой как Remez или Meteor.

При реализации измерителя пиков также может предъявляться требование исключения воздействия любого внешнего источника постоянного тока, поскольку традиционно измерители звуковых сигналов заперты по постоянному току. С другой стороны, если значение истинного пика сигнала

требуется в целях исключения перегрузки, тогда присутствие постоянного тока должно сохраняться и измеряться. При необходимости исключение постоянного тока можно выполнить, задействовав малую вычислительную мощность, путем включения фильтра верхних частот низшего порядка с бесконечной импульсной характеристикой (ИИХ) на входе измерителя.

Иногда необходимо измерять амплитуду пикового сигнала после применения определенного типа взвешивающего фильтра, с тем чтобы выделить воздействие некоторых участков полосы частот. Реализация зависит от свойств конкретного взвешивающего фильтра.
