

МЕТОД КОЛИЧЕСТВЕННОЙ ОЦЕНКИ СЛУХОВОГО ВОСПРИЯТИЯ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ. ЧАСТЬ 1. ИСПЫТАТЕЛЬНЫЕ СИГНАЛЫ И МЕРЫ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ

М.О. Пахомов^{1*}, В.С. Рожнов^{1,2}

¹Санкт-Петербургский филиал корпорации LG Electronics inc.,

Санкт-Петербург, 191123, Российская Федерация

²Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,

Санкт-Петербург, 193232, Российская Федерация

*Адрес для переписки: mpakhomov39@gmail.com

Информация о статье

УДК 621.391.832.43

Язык статьи – русский

Ссылка для цитирования: Пахомов М.О., Рожнов В.С. Метод количественной оценки слухового восприятия нелинейных искажений. Часть 1. Испытательные сигналы и меры количественной оценки нелинейных искажений // Труды учебных заведений связи. 2018. Т. 4. № 1. С. 75–84.

Аннотация: Изложена процедура выделения сигнала нелинейных искажений, возникающих в звуковых трактах различных устройств, при одновременном наличии в воспроизводимом сигнале частотных и динамических искажений. Предложены критерии оценки их слуховой заметности, и весовые коэффициенты, учитывающие распределение энергии сигнала нелинейных искажений по частоте и длительности заметно искаженных фрагментов. Получены предварительные экспериментальные данные, подтверждающие зависимость слухового восприятия нелинейных искажений в зависимости от тональной структуры сигнала и спектрального распределения отношения «Шум-Маска»

Ключевые слова: многотональные испытательные сигналы, нелинейные искажения, одновременная маскировка, отношение «Шум-Маска».

1. ВВЕДЕНИЕ

Существующие методы оценки нелинейных искажений можно разделить на несколько групп с точки зрения учета, как характеристик выбранного испытательного сигнала, так и закономерностей слухового восприятия. Стандартные метрики количественной оценки нелинейных искажений (коэффициент гармонических искажений, коэффициент интермодуляций), а также методы использующие ряды Вольтера для расчета гармоник высоких порядков не имеют явной привязки к характеристикам испытательного сигнала и учитывают такого важного феномена слухового восприятия как маскировка. В предлагаемой работе оба этих недостатка устранены. Для точного выделения и оценки слуховой заметности нелинейных искажений из сильно искаженного музыкального сигнала, содержащего одновременно различные виды искажений при значительной величине каждого из них, предлагается использовать синтези-

рованный многотональный сигнал, амплитудно-частотно-временные характеристики которого соответствуют по целому ряду параметров музыкальному сигналу, используемому при проведении экспертиз. При этом для получения оценок, совпадающих с данными субъективно-статистических экспертиз, необходимо учитывать эффекты маскировки полезным сигналом нелинейных искажений, динамические, временные и частотные особенности последнего, существенно влияющие на его слуховую заметность.

2. ПОСТАНОВКА ЗАДАЧИ

В [1], [2] исследовано влияние нелинейных искажений на субъективное восприятие музыкальных и речевых сигналов. В качестве испытательного сигнала автором этой работы использован синтезированный многотональный сигнал, предложенный ранее в [3], [4]. Подобные многотональные сигналы с различным распределением спектраль-

ных компонент по частоте и постоянным уровнем каждой спектральной компоненты подробно описаны в [5]. На наш взгляд, у таких сигналов, содержащих определенное число тональных компонент постоянной амплитуды, имеется существенный недостаток – отсутствие связи со структурой реальных музыкальных сигналов, используемыми экспертами при тестовом прослушивании.

Известно, что величина и слуховая заметность нелинейных искажений, вносимых в сигнал звуковыми трактами испытываемых устройств, зависит от амплитудно-частотно-временных характеристик самого испытательного сигнала [6]. При этом для обеспечения высокой корреляции между экспертными оценками и данными измерений необходимо выполнить два условия:

- обеспечить точное выделение сигнала нелинейных искажений, внесенных в испытательный сигнал оцениваемым звуковым трактом;

- устранить энергетические, частотные и временные различия между испытательным и соответствующим ему музыкальным сигналом в каждый текущий момент времени.

При использовании в качестве испытательного сигнала музыкальных фонограмм наибольшую сложность представляет задача выделения сигнала нелинейных искажений из исходного музыкального сигнала, прошедшего через звуковой тракт испытываемого устройства. В психоакустической модели PEAQ (*от англ.* Perceptual Evaluation of Audio Quality) [7], [8] для этого используется вычитание спектров референсного музыкального и тестового сигналов с последующей компенсацией частотных искажений, возникающих в звуковых трактах одновременно с нелинейными. Однако при одновременном присутствии в оцениваемом сигнале частотных и динамических искажений точное выделение сигнала нелинейных искажений является крайне сложной задачей.

В данной работе изложен метод, позволяющий, с одной стороны, обеспечить точное выделение сигнала нелинейных искажений из музыкального сигнала, прошедшего звуковой тракт испытываемого устройства и, с другой стороны, получить при этом для него близкие к референсному музыкальному сигналу амплитудно-частотно-временные характеристики. Иными словами, ниже изложен метод синтеза многотонального сигнала, спектр которого имеет идентичное исходному музыкальному сигналу частотно-временное распределение энергии в любой текущий момент времени.

Для большей строгости последующего изложения введем следующие пояснения используемых терминов:

- референсный музыкальный сигнал – музыкальный сигнал, отобранный экспертами для оценки качества звуковых трактов испытываемых устройств;

- референсный многотональный сигнал – многотональный сигнал, синтезированный из референсного музыкального сигнала;

- тестовый музыкальный сигнал – референсный музыкальный сигнал, прошедший через звуковой тракт испытываемого устройства и записанный с помощью микрофона;

- тестовый многотональный сигнал – референсный многотональный сигнал, прошедший через звуковой тракт испытываемого устройства, и также записанный с помощью микрофона;

- очищенный тестовый многотональный сигнал – сигнал, имеющий в своем составе только спектральные компоненты тестового многотонального сигнала, содержащиеся в референсном многотональном сигнале;

- сигнал нелинейных искажений – сигнал, содержащий только новые спектральные компоненты тестового многотонального сигнала после его прохождения через звуковой тракт испытываемого устройства, т.е. изначально отсутствующие в референсном многотональном сигнале.

3. МНОГОТОНАЛЬНЫЙ СИГНАЛ

В рамках данной работы предложено использовать три типа структур референсных многотональных сигналов, применяемых для оценки величины нелинейных искажений, возникающих в звуковом оборудовании различного назначения: стационарный, квазистационарный и прореженный музыкальный. Перечислим кратко основные достоинства и недостатки каждого из них.

3.1. Стационарный и квазистационарный референсные многотональные сигналы

Стационарный референсный многотональный сигнал содержит наборы тональных компонент определенной частоты, амплитуды и фазы, квазистационарный – тональные компоненты с изменяющейся во времени частотой, амплитудой и фазой, отражающей текущие изменения в референсном музыкальном сигнале.

Стационарный сигнал содержит набор тональных компонент, расположенных в центрах $1/3$ октавных частотных полос (всего 24 компоненты). Его длительность выбирается равной длительности референсного музыкального сигнала. Оба сигнала разделяются на выборки длиной 1024 отчета с 50% перекрытием. Для каждой выборки с помощью быстрого преобразования Фурье (БПФ) вычисляется усредненная в $1/3$ октавных полосах частот энергия каждого из них, при этом энергия соответствующих компонент созданного референсного многотонального сигнала устанавливается равной значениям энергии исходного референсного музыкального сигнала в соответствующих полосах частот (рисунок 1).

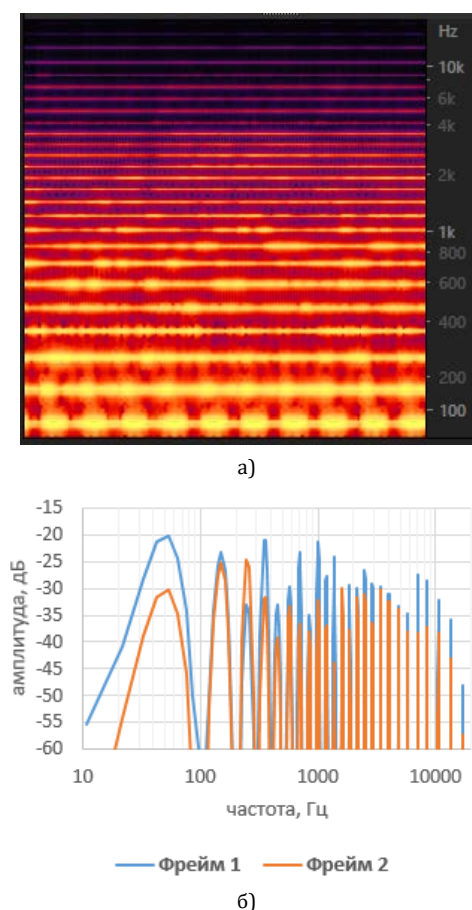


Рис. 1. Стационарный многотональный сигнал:
а) спектрограмма; б) спектры произвольных выборок

Спектрограммы показывает, что частоты компонент не меняются на всей длительности сигнала, а меняются только амплитуды.

Достоинством данного подхода является простота создания стационарного референсного многотонального сигнала, выделения сигнала нелинейных искажений и расчета порога маскировки при оценке слуховой заметности нелинейных искажений. Недостатком такого тестового сигнала является неполное соответствие характеристик стационарного многотонального сигнала структуре исходного музыкального сигнала. Это может приводить к существенному расхождению характеристик сигналов нелинейных искажений, выделяемых из соответствующих тестовых сигналов (референсных музыкального и многотональный сигналов, пропущенных через испытуемый звуковой тракт).

Структура квазистационарного референсного многотонального сигнала также содержит 24 тональных компоненты, однако их частоты, амплитуды и фазы не постоянны на всем протяжении сигнала – в каждый текущий момент времени они выбираются в соответствии с параметрами максимальных спектральных компонент соответствующей выборки исходного сигнала. Для того, чтобы избежать появления дополнительных ис-

кажений, возникающих при выполнении прямого БПФ (известных как эффект Гиббса), выборка референсного музыкального сигнала предварительно взвешивается окном Ханна, при этом для каждой одной выборки референсного музыкального сигнала создаются 3 выборки референсного многотонального сигнала с 50 % перекрытием. Частота спектральной компоненты в каждой 1/3 октавной полосе частот создаваемого референсного многотонального сигнала, соответствует частоте максимальной компоненты референсного музыкального сигнала из соответствующей выборки, а ее уровень устанавливается равным уровню энергии в 1/3 октавной полосе исходного сигнала. Фазы компонент при этом выбираются случайным образом. В случае отсутствия тональных компонент в некоторой 1/3 октавной полосе референсного музыкального сигнала в центр соответствующей частотной полосы референсного многотонального сигнала добавляется тональная компонента с энергией, равной суммарной энергии шумовых компонент исходного референсного музыкального сигнала в данной полосе. В результате для каждой его выборки длиной 1024 отсчета создается набор стационарных фрагментов референсного многотонального сигнала длительностью $1024 \cdot 2 = 2048$ отсчетов сигнала каждый, по три фрагмента с 50 % перекрытием (рисунок 2).

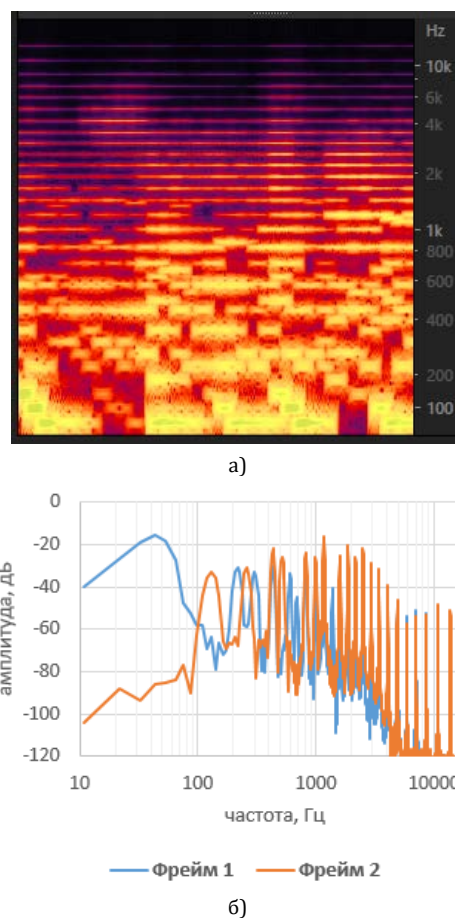


Рис. 2. Квазистационарный многотональный сигнал:
а) спектрограмма; б) спектры произвольных выборок

Спектрограмма показывает, что отдельные фрагменты сигнала отличаются друг от друга не только амплитудами компонент, но и их частотой. На высоких частотах спектр сигнала состоит из набора тонов с фиксированной частотой, поскольку в исходном музыкальном сигнале в этих полосах частот не нашлось компонент соответствующих критерию отбора.

Для каждого фрагмента создается запись в паспорте референсного многотонального сигнала, содержащая информацию о типе спектральной компоненты – тональная или шумовая (рисунок 3).

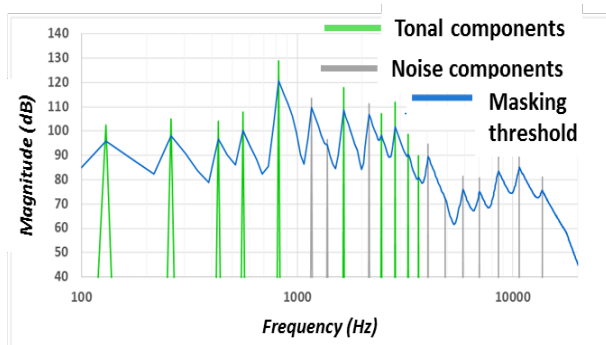


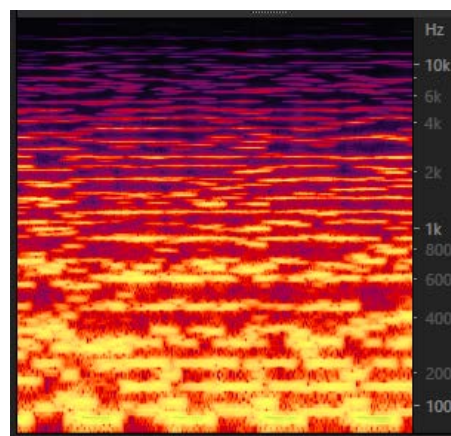
Рис. 3. Спектр выборки квазистационарного сигнала с разделением на тональные и шумовые компоненты

Достоинством второго подхода является более полное соответствие квазистационарного референсного многотонального сигнала структуре исходного референсного музыкального сигнала, чем в первом случае. Недостатком может служить нарушение фазовой характеристики сигнала, увеличенная в два раза длительность и замена реальных шумовых компонент тональными, что приводит к появлению дополнительных интермодуляционных искажений. Поскольку наиболее критичными с точки зрения возникновения нелинейных искажений являются участки сигнала, имеющие максимальную амплитуду и частоту, близкую к частоте резонанса громкоговорителя, то имеет смысл создавать референсный многотональный сигнал не по всей длительности референсного музыкального сигнала, а только по его фрагментам, относящимся к указанным выше отрезкам.

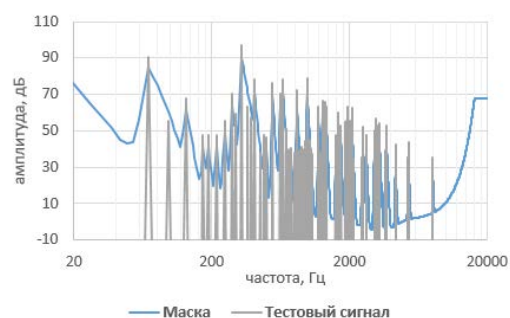
3.2. Прореженный по частоте референсный многотональный сигнал

Данный вариант референсного многотонального сигнала создается путем удаления из исходного референсного музыкального сигнала спектральных компонент с малой энергией. Известны два варианта получения сигнала, различающиеся степенью прореживания спектра референсного музыкального сигнала. При процедуре прореживания сигнал разделяется на выборки длиной 2048 отсчетов с 50 % перекрытием, далее с помощью БПФ вычисляется спектр каждой из них. Спектр полученного сигнала разделяется на полосы, соответствующие критическим полосам слуха, и в

каждой такой полосе, в зависимости от варианта реализации, выбирается либо одна максимальная спектральная компонента, либо несколько компонент, соответствующих некоторому критерию отбора (например, уровень локального максимума должен превышать соседние локальные минимумы на 7 дБ); прочие спектральные компоненты обнуляются. В случае отсутствия явно выраженных максимумов сигнала в некоторой полосе, имеющийся шумовой сигнал заменяется тональной компонентой, расположенной в центре этой полосы и с такой же энергией. Далее производится подстройка уровней выбранных спектральных компонент для уравнивания энергий исходного и прореженного сигналов в критических полосах слуха. В заключении осуществляется перевод полученного сигнала во временную область и проверка его максимальных энергий в выбранных частотных полосах. Для каждой выборки сигнала создается запись в паспорте, содержащая информацию о наличии и типе компоненты в каждой критической полосе (рисунок 4).



а)



б)

Рис. 4. Прореженный музыкальный сигнал: а) спектрограмма; б) спектр произвольной выборки и глобальный порог маскировки

Из спектрограммы видно, что в отличие от стационарного и квазистационарного искусственных многотональных сигналов спектральная структура прореженного музыкального сигнала наиболее близка структуре исходного референсного музыкального сигнала в каждый текущий момент времени. Это является достоинством такого подхода,

но и определяет его основной недостаток – сложность выделения сигнала нелинейных искажений.

Следует отметить, что для последующей синхронизации тестовых и референсных сигналов, в начало и в конец каждого из них добавляются специальные последовательности тональных импульсов.

4. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

В процессе исследований в испытательный сигнал вносились нелинейные искажения, как в отдельных узких полосах частот, так и одновременно во всем диапазоне слышимых частот. Уровень сигнала нелинейных искажений был привязан к порогу его маскировки полезным сигналом в каждом диапазоне частот.

4.1. Вносимые нелинейные искажения

Для внесения нелинейных искажений в референсные музыкальные и многотональные сигналы использовались способы, подробно описанные в [1], а именно: мягкое и жесткое, симметричное и асимметричное клиппирование сигнала в выбранной полосе частот, смешивание испытательных сигналов с белым шумом, пропускание сигналов через звуковые тракты, обеспечивающие внесение разных видов нелинейных искажений в испытательный сигнал.

Исследовалось слуховое восприятие следующих типов искажений:

- жесткое симметричное ограничение сигнала с различными уровнями и длительностями существования, при этом уровни ограничения выбирались на основе распределений мгновенных значений амплитуд РМС;
- жесткое асимметричное ограничение положительных пиков сигнала во временной области (с различными уровнями ограничения);
- мягкое симметричное ограничение с разными параметрами компрессора;
- мягкое асимметричное ограничение положительных пиков сигнала во временной области с разными параметрами компрессора;
- центрирование, при котором отсчеты сигнала в пределах определенного диапазона изменения амплитуд устанавливались равными 0 (диапазоны ограничения были установлены в % от среднеквадратичного значения входного сигнала);
- пропускание сигнала через звуковые тракты с различными уровнями гармонических искажений (величины гармонических искажений для 10 % значений мгновенных амплитуд референсного музыкального сигнала составляли от 5 до 50 % в разных диапазонах частот).

При моделировании узкополосных нелинейных искажений референсные музыкальные сигналы обрабатывались соответствующими фильтрами для выделения низкочастотной части сигнала, которая подавалась на вход компрессоров. Полу-

ченные после цифровой фильтрации сигналы в полосе частот выше частоты среза фильтра содержали только компоненты искажений. С помощью полосовых фильтров формировался набор сигналов нелинейных искажений в $\frac{1}{2}$ октавных полосах частот. Далее полученные сигналы объединялись и микшировались с референсными музыкальными сигналами.

В результате был получен широкий набор сигналов нелинейных искажений разных типов с различным амплитудно-частотно-временным распределением. Для чего было создано специальное программное обеспечение, которое также используют для изменения уровня (дозирования) искажений в разных полосах частот.

4.2. Эксперты

В качестве экспертов привлекались молодые люди (возраст от 24 до 31 года, 8 человек) без дефектов слуха, имеющие аудиометрический порог ниже 20 дБ в полосе частот 125 Гц – 8 кГц. Для оценки слуховой заметности вносимых искажений использовалась семибалльная шкала, при этом наиболее предпочтительное звучание получало оценку 7 баллов, наименее предпочтительное – 1.

4.3. Тестовые сигналы

Референсные музыкальные сигналы представляли собой звуковые отрывки различных жанров – джаз, поп, рок музыка, женский вокал с аккомпанементом, фортепьяно, классическая музыка в исполнении оркестра. Все записи были взяты с коммерческих компакт-дисков. Выбранные отрывки имели длительность от 10 до 15 с.

Референсные музыкальные сигналы обрабатывались созданным программным обеспечением (для внесения требуемых нелинейных искажений); они также подавались на звуковые тракты испытываемых устройств. Запись сигналов осуществлялась в звукоизолированной камере с уровнем фонового шума в полосе частот от 100 Гц до 20 кГц ниже 20 дБ. Далее эти тестовые сигналы обрабатывались фильтрами для устранения внесенных частотных искажений и затем предлагались экспертам для получения субъективной оценки. Прослушивания проводились в звукоизолированной комнате. Тестовые музыкальные сигналы через внешнюю 24-х битовую звуковую карту MOTU UltraLite-mk3 подавались на усилитель HiFiMan EF-100, после чего воспроизводились с помощью высококачественных головных телефонов Sennheiser HD 800, имеющих крайне низкий коэффициент нелинейных искажений ($< 1\%$).

Перед началом экспериментальных исследований экспертам были предъявлены образцы отобранных референсных музыкальных сигналов (свободных от искажений, с оценкой 7 баллов). Аналогичным образом были предъявлены также

образцы сигналов с максимальным количеством внесенных нелинейных искажений (как наихудшее звучание, с оценкой 1 балл). Экспертам предлагалось прослушать весь набор записей в любом произвольном порядке для оценки их качества, затем, прослушивая повторно, поставить оценку в соответствии с собственным предпочтением, используя семибалльную шкалу. Оценки различных экспертов усреднялись. Доверительный интервал для оценок экспертов составил 0,54 балла.

5. ВЫДЕЛЕНИЕ СИГНАЛА НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ

Тестовые многотональные сигналы, полученные программным путем или прошедшие звуковой тракт испытуемого устройства, помимо полезного сигнала содержат в себе также сигнал частотных, нелинейных и динамических искажений. Обычно энергия сигнала частотных искажений на порядок превосходит энергию сигнала нелинейных искажений.

Как было отмечено ранее, для оценки слуховой заметности нелинейных искажений и их влияния на субъективное предпочтение необходимо обеспечить точное выделение сигнала нелинейных искажений. Обычно эта задача решается с помощью свертки тестового сигнала с инвертированной импульсной характеристикой оцениваемого звукового тракта, последующей нормализацией и вычислением разницы спектров полученного сигнала и референсного сигнала. Однако в случае присутствия в испытуемом устройстве значительных динамических искажений указанный способ не позволяет достичь необходимой точности при выделении сигнала нелинейных искажений, так как импульсная характеристика тракта изменяется в момент срабатывания ограничителя.

Приведем краткое описание процедуры выделения сигнала нелинейных искажений из пары референсного и тестового многотональных сигналов. Предварительно оба сигнала синхронизируются с точностью до < 5 отсчетов, затем разбиваются на выборки, после чего взвешиваются оконной функцией и с помощью БПФ переводятся в частотную область. Следует отметить, что алгоритмы выделения сигнала нелинейных искажений имеют отличия в зависимости от версии многотонального сигнала. Далее спектр выборки тестового многотонального сигнала разделяется на два набора компонент – первый включает только те компоненты, которые содержатся в спектре соответствующей выборки референсного многотонального сигнала. Эти компоненты образуют очищенный тестовый многотональный сигнал. Второй набор содержит новые компоненты, отсутствующие в спектре исходного многотонального сигнала – они представляют собой сигнал нелинейных искажений (рисунок 5).

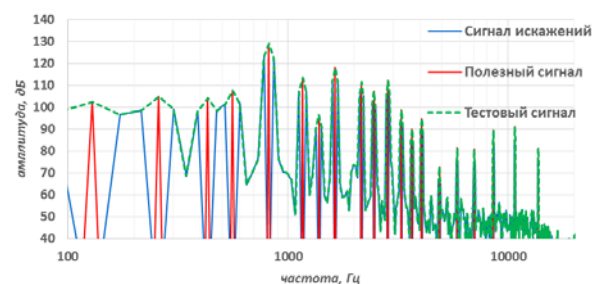


Рис. 5. Спектр выборки тестового сигнала с разделением его на полезный сигнал и сигнал нелинейных искажений

6. ОЦЕНКА СЛУХОВОЙ ЗАМЕТНОСТИ СИГНАЛА НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ

Для оценки влияния сигнала нелинейных искажений на субъективную оценку качества необходимо учитывать эффект маскировки полезным сигналом и разную чувствительность слуховой системы человека в различных диапазонах частот [9], [10].

Для учета эффекта маскировки сигнала нелинейных искажений полезным сигналом использовались разные психоакустические модели. Для стационарного и квазистационарного сигналов использовалась упрощенная психоакустическая модель стандарта MPEG-1 ISO/IEC 11172-3.

Анализируемый тестовый многотональный сигнал разбивается на выборки, длительность которых зависит от типа используемого многотонального сигнала и для каждой выборки сигнала вычисляется энергетический линейчатый спектр $X(k)$. Для уменьшения краевых эффектов при выполнении БПФ используется окно Ханна $h(n)$. Для полученного энергетического спектра выполняется процедура нормирования, здесь максимальной спектральной компоненте присваивается уровень 96 дБ, а все значения остальных спектральных компонент подвергаются процедуре коррекции относительно этой величины. Для каждой тональной и шумоподобной компонент (информация о типе компоненты берется из паспорта сигнала) выполняется расчет коэффициентов маскировки и индивидуальных кривых маскировки $M[z(i), z(j)]$, и затем для них вычисляются пороги маскировки $N_{тк}[z(i)]$ и $N_{шк}[z(i)]$. В завершении, рассчитанные пороги суммируются вместе с абсолютным порогом слышимости для получения глобального порога маскировки $N_{пм}(i)$ в каждой субполосе психоакустического анализа [11], [12], [13].

Следует отметить, что аппроксимации, приведенные в разных источниках, могут недостаточно соответствовать реальным порогам маскировки для конкретного слушателя. Поэтому коэффициенты и форма порогов редактировалась в процессе проведения тестов прослушивания таким образом, чтобы продукты нелинейных искажений, регистрируемые слушателями с вероятностью выше 75 %, гарантированно превышали порог маскировки.

При использовании в качестве испытательных сигналов прореженных музыкальных сигналов было принято решение использовать стандарты MPEG-2 ISO/IEC 13818-7 и MPEG-2 ISO/IEC 14496-3, поскольку они обеспечивают более точное вычисление порога маскировки.

Анализируемый сигнал разбивается на выборки длиной 1024 отсчета, с помощью БПФ вычисляется комплексный спектр исходной выборки и, как и в первом случае, для взвешивания отсчетов, используется окно Ханна $h(n)$. Далее осуществляется предсказание текущих значений амплитудных и фазовых значений спектральных компонент на основе информации, полученной из двух предыдущих выборок и вычисляется так называемая мера непредсказуемости C_w (от англ. unpredictability) для каждой спектральной компоненты текущей выборки. На основе этого критерия принимается решение о корреляции между спектральными компонентами текущей и двух предыдущих выборок звукового сигнала. На следующем шаге вычисляются энергия e_b и взвешенное значение меры непредсказуемости c_b текущей выборки в каждой полосе психоакустического анализа b .

Затем рассчитывается так называемая развертывающая функция $M(i, j)$, представляющая собой индивидуальную кривую маскировки, учитывающую избирательные свойства базилярной мембраны уха, и вычисляются ее свертка с функциями энергии сигнала и взвешенными значениями меры непредсказуемости с целью учета влияния соседних полос психоакустического анализа. В завершении выполняется расчет индексов тональности α_b . Более подробно используемые алгоритмы описаны в [14].

Далее, рассчитанный порог маскировки M сравнивался с распределением энергии E_{noise} , соответствующей выборки ШНИ по частоте, и рассчитывалось отношение «Шум-Маска» (NMR, от англ. Noise-to-Mask-Ratio) в диапазоне частот по формуле:

$$NMR_z(n) = \frac{1}{(Z2 - Z1 + 1)} \sum_{k=Z1}^{Z2} \frac{E_{noise}[k, n]}{M[k, n]}, \quad (1)$$

где $Z1$ – начальная дискрета учитываемого диапазона частот; $Z2$ – конечная дискрета учитываемого диапазона частот; n – номер выборки; z – номер диапазона частот.

В существующих психоакустических моделях величина порога одновременной маскировки (NMR) рассчитывается для каждой отдельной выборки с последующим усреднением по всему сигналу [7], [15]. Однако, как показывают эксперименты, музыкальные сигналы с внесенными сигналами нелинейных искажений с одинаковым значением NMR при прослушивании могут иметь разную оценку качества с точки зрения предпочтения.

В нашем случае предлагается вычислять величину, соответствующую слуховому восприятию нелинейных искажений, в зависимости от распределения значений отношения энергии шумов к порогу маскировки по частоте: PNL (от англ. Perceived Noise Level). Величина PNL рассчитывается, как взвешенная сумма величин NMR, рассчитанных по формуле (1) в октавных полосах частот j :

$$PNL(n) = \frac{1}{8} \sum_{j=1}^8 NMR_j(n) \cdot G(j), \quad (2)$$

где n – номер выборки; j – номер октавной полосы частот. Значения весовых коэффициентов $G[j]$ в каждой полосе частот j определяется на основе анализа данных субъективных экспертиз и результатов анализа распределения величин NMR по частоте методом множественной линейной регрессии.

Влияние нелинейных искажений на субъективную оценку качества воспроизводимого испытуемым устройством сигнала определяется средней величиной PNL, полученной для всего сигнала в целом:

$$PNL = \frac{1}{N} \sum_{i=1}^N PNL(i), \quad (3)$$

где N – количество выборок в сигнале, i – порядковый номер выборки.

Для удобства приведения вычисленной меры нелинейных искажений к субъективной оценке экспертов предлагается разделять шкалу значений PNL на уровни, соответствующие достоверному разделению субъективных оценок предпочтения звучания искаженных музыкальных сигналов при тестах прослушивания. Следует отметить, что влияние на слуховую заметность сигнала нелинейных искажений оказывает не только средний уровень PNL для звукового сигнала в целом, но и распределение этой величины как по частоте, так и во временной области (рисунок 6).

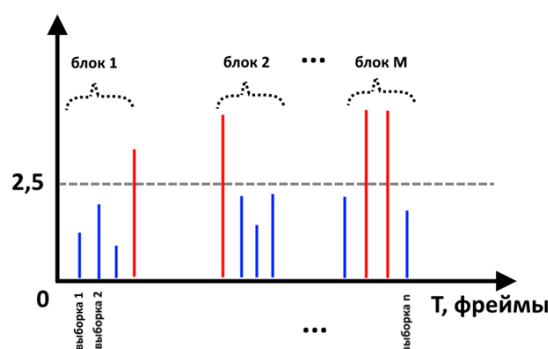


Рис. 6. Пример распределения метрики PNL по длительности сигнала

Влияние сигнала нелинейных искажений на субъективную оценку качества при воспроизведении определяется как относительным количеством сильно искаженных выборок – RDF (от англ. Related Distorted Frames):

$$RDF = \frac{1}{N} \sum_{k=1}^N \begin{pmatrix} 1, PNL(i) > 2,5dB \\ 0, PNL(i) \leq 2,5dB \end{pmatrix}, \quad (4)$$

так и количеством блоков искаженных выборок M и средней длительностью каждого такого блока – ALDB (*от англ.* Average Length of Distorted Blocks):

$$ALDB = \frac{1}{M} \left[\sum_{m=1}^M LDB_m \right], \quad (5)$$

где LDB_m (*от англ.* Length of Distorted Block) – длительность m -го блока искаженных выборок.

7. ЭКСПЕРИМЕНТЫ

Для определения зависимости взвешивающей функции $G(j)$ от частоты были проведены две серии субъективно-статистических экспертиз, для которых был подготовлен набор сигналов с известным количеством нелинейных искажений, присутствующих как в отдельных полосах частот, так и в нескольких частотных диапазонах одновременно.

В первой серии прослушиваний экспертам предлагалось оценить с точки зрения предпочтительности набор тестовых музыкальных сигналов, имеющих нелинейные искажения только в одной полосе частот. Всего было предложено для прослушивания 62 тестовых фонограммы; экспертам требовалось рассортировать их на 7 групп по степени предпочтительности. С помощью этой серии экспериментов удалось определить индивидуальные пороги одновременной маскировки для отдельных экспертов и зависимости субъективной оценки от величины NMR в полосе частот, содержащей нелинейные искажения (рисунок 7). Вычисление NMR проходило на основе использования прореженного музыкального сигнала и психоакустической модели AAC (*от англ.* Advanced Audio Codec).

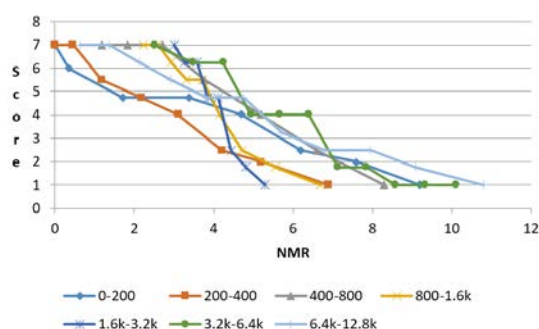


Рис. 7. Кривые изменения предпочтительности звучания от величины отношения «Шум-Маска» при наличии нелинейных искажений только в одной полосе частот

Из графиков следует, что скорость деградации оценки субъективного предпочтения зависит от диапазона частот, в котором присутствуют слышимые нелинейные искажения. Также следует отметить, что в зависимости от диапазона частот некоторое количество слышимых искажений мо-

жет не приводить к снижению рейтинга предпочтения фонограммы.

Во второй серии прослушиваний экспертам предлагалось ранжировать с точки зрения предпочтительности набор тестовых фонограмм, имеющих слышимые нелинейные искажения одновременно в нескольких полосах частот. Для проведения этой серии прослушиваний были сформированы два набора референсных музыкальных сигналов, каждый из которых содержал около 200 образцов. Каждый образец представлял собой фрагмент музыкальной композиции, смикшированный с сигналами нелинейных искажений, содержащих различные комбинации уровня искажений в разных диапазонах частот. Результаты обработки второй серии тестов для каждого набора сведены в таблицы, подобные таблице 1.

ТАБЛИЦА 1. Обработка результатов экспертных прослушиваний

Фонограмма	Усредненные оценки	Величина NMR в диапазоне				
		Band 1	Band 2	...	Band8	
1	Y_1	$X_{1,1}$	$X_{1,2}$			$X_{1,8}$
2	Y_2	$X_{2,1}$	$X_{2,2}$			$X_{2,8}$
...						
K	Y_K	$X_{K,1}$	$X_{K,2}$			$X_{K,8}$

Данная таблица представляет собой по сути де-ла массив исходных данных для получения весовых коэффициентов $G(j)$, оценивающих влияние нелинейных искажений в том или ином диапазоне частот на субъективную оценку предпочтения воспроизводимого сигнала.

Определение $G(j)$ осуществляется с помощью метода множественной линейной регрессии:

$$y_i = \sum_{j=1}^8 G_j \cdot x_{ij} + \varepsilon_i = XW + \varepsilon, \quad i = 1, \dots, K, \quad (6)$$

где $Y(y_1, \dots, y_i, \dots, y_n)$ – вектор усредненных субъективных оценок предпочтения; $X = (8 \times K)$ – матрица величин отношения «Шум-Маска» (строки матрицы – векторы значений величин NMR в частотных диапазонах для i -ой тестовой фонограммы); x_{ij} – значения NMR в j -ом диапазоне; K – число тестовых фонограмм; ε – случайная составляющая; G_j – коэффициенты регрессии (искомые весовые коэффициенты).

Решение системы (6) осуществлялось с помощью метода наименьших квадратов.

В тестах использовались две музыкальные фонограммы: первая – фрагмент жанра Pop-Music, вторая – фрагмент вокальной партии. Фрагменты имели сильное различие с точки зрения коэффициента тональности.

Коэффициент тональности [14] для фрагмента вычислялся на основе вычисления меры пологости спектра в полосе частот:

$$Tonality = \min\left(\frac{SFM_{db}}{-60}, 1\right), \quad (7)$$

где $SFM_{db} = 10 \cdot \log_{10}(SFM)$; SFM (от англ. Spectral Flatness Measure) – мера спектральной пологости, вычисляемая по формуле:

$$SFM(num_{band}) = \frac{\left(\prod_{k \in num_{band}} a(k)\right)^{\frac{1}{K}}}{\frac{1}{K} \sum_{k \in num_{band}} a(k)}, \quad (8)$$

где $a(k)$ – амплитуды k -ой компоненты спектра.

Для вычисления индекса тональности использовались 4 октавных диапазона частот от 250 Гц до 4000 Гц.

Для первой композиции усредненный коэффициент тональности равен 0.047, для второй – 0.24 соответственно. Результаты вычисления взвешивающей функции $G(j)$ приведены на рисунке 8.



Рис. 8. Зависимости взвешивающей функции G от полосы частот и коэффициента тональности сигнала.

Проведенные эксперименты показывают, что звуковые сигналы с вокальными частями с высоким коэффициентом тональности очень чувствительны к искажениям в голосовом диапазоне частот, тогда как сигналы с низким уровнем тональности, напротив, более чувствительны к звуковым искажениям в высокочастотной области, что, возможно, связано с увеличением меры резкости сигнала.

8. ЗАКЛЮЧЕНИЕ

В работе представлена новая модель оценки слуховой заметности нелинейных искажений и метод прогнозирования субъективной оценки предпочтительности музыкальных сигналов, подвержен-

ных нелинейным искажениям. Модель описывает процедуры создания искусственных тестовых сигналов, а метод – извлечение сигнала нелинейного искажения из тестового сигнала, пропущенного через звуковой тракт. При оценке слуховой заметности нелинейных искажений он учитывает эффект одновременной маскировки сигнала нелинейных искажений полезным сигналом, рассчитывает значение отношения «Шум-Маска» для выбора сигнала, учитывает зависимость этого отношения от частоты с помощью весовой функции.

Достоверность объективных оценок, полученных с помощью предложенного метода, проверена статистическими методами. Коэффициент корреляции Пирсона для серии прогнозируемых и полученных в результате тестов прослушивания оценок составил 0,96 для музыкальных сигналов с жестким и мягким ограничением. При передаче музыкальных сигналов через реальные звуковые тракты с различным уровнем нелинейных искажений коэффициент Пирсона снизился до 0,84 для метода, основанного на применении стационарного и квазистационарного многотонального сигнала, и до 0,9 – для метода, основанного на применении «прореженного» тестового музыкального сигнала. Вероятной причиной уменьшения корреляции объективных и субъективных оценок является отсутствие учета временного распределения искаженных выборок в случае небольшого количества нелинейных искажений.

Основными направлениями дальнейших исследований, результаты которых планируются к публикации во второй части работы, являются:

- определение весовых коэффициентов вклада нелинейных искажений в отдельных полосах частот с учетом длительности их непрерывного существования;
- исследование зависимости весовой функции $G(j)$ от распределения спектральной плотности мощности испытательного сигнала по частоте;
- оценка влияния коротких и редких искаженных участков сигнала на оценку качества.

Экспериментальное изучение перечисленных выше факторов позволит разработать метод надежного прогнозирования субъективных оценок предпочтения для звуковых трактов с различными видами нелинейностей.

Благодарности

Авторы выражают благодарность Ковалгину Юрию Алексеевичу – доктору технических наук, профессору, заслуженному деятелю науки России, профессору кафедры радиосвязи и вещания Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, за научное рецензирование работы.

Список используемых источников

1. Tan C.T., Moore B. C. J., Zacharov N. The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals. J. Audio Eng. Soc. Vol. 51. PP. 1012–1031 (2003 Nov.).
2. Tan C.T., Moore B.C. J., Zacharov N., Mattila V.V. Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals. J. Audio Eng. Soc. Vol. 52. PP. 699–711 (2004 July/Aug.).

3. Czerwinski E., Voishvillo A., Alexandrov S. and Terekhov A., Multitone Testing of Sound System Components–Some Results and Conclusions. Part 1: History and Theory. J. Audio Eng. Soc. Vol 49. PP. 1011–1042 (2001 Nov.).
4. Czerwinski E., Voishvillo A., Alexandrov S. and Terekhov A. Multitone Testing of Sound System Components–Some Results and Conclusions. Part 2: Modeling and Application. J. Audio Eng. Soc. Vol 49. PP. 1181–1192 (2001 Dec.).
5. Risch J.M. A New Class of In-Band Multitone Test Signals. Presented at the 105th Convention of the Audio Engineering Society. J. Audio Eng. Soc. (Abstracts). Vol. 46. P. 1037 (1998 Nov.), preprint 4803.
6. Zwicker E. Psychoakustik // Berlin – Heidelberg – New York: Springer Verlag. 1982.
7. ITU-R Recommendation BS.1387-1, Method for Objective Measurements of Perceived Audio Quality. Rec. ITU-R, (1998-2001).
8. Thiede T. PEAQ – The ITU Standard for Objective Measurement of Perceived Audio Quality // J. Audio Eng. Soc. Jan.–Feb. 2000. Vol. 48. PP. 3–29.
9. Fastl H. Temporal Masking Effects: II. Critical Band Noise Masker / Acustica. 1976. Vol. 36. PP. 317–331.
10. Fastl H., Zwicker E. Psychoacoustics: Facts and Models // 3rd Edition. Springer–Verlag. 2007.
11. Hellman R.P. Asymmetry of Masking between Noise and Tone // Perception & Psychophysics. 1972. Vol. 11 (3). PP. 241–246.
12. Humes L.E. Models of the Additivity of Masking // Journal of the Acoustical Society of America. 1989. Vol. 85 (3). PP. 1285–1294.
13. Moore B.C.J., Glasberg B.R., Baer T. A Model for the Prediction of Thresholds, Loudness, and Partial Loudness. J. Audio Eng. Soc. Vol. 45. PP. 224–240 (1997 Apr.).
14. Ковалгин Ю.А., Вологдин Э.И. Цифровое кодирование звуковых сигналов: Учебное пособие. СПб.: КОРОНА-принт, 2004. 240 с.
15. Borrás J., Hatrack P., Mandayam N. Decision Theoretic Framework for NLOS Identification // Vehicular Technology Conference. VTC 98. 48th IEEE, Ottawa, Ont. 1998. Vol. 2. PP. 1583–1587.

* * *

METHOD FOR QUANTITATIVE EVALUATION OF AUDITORY PERCEPTION OF NONLINEAR DISTORTION. PART I. TEST SIGNALS & MEASURE OF QUANTIFICATION

M. Pakhomov¹, V. Rozhnov^{1,2}

¹St. Petersburg Branch of the corporation LG Electronics inc.
St. Petersburg, 191123, Russian Federation

²The Bonch-Bruевич State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Article info

Article in Russian

For citation: Pakhomov M., Rozhnov V. Method for Quantitative Evaluation of Auditory Perception of Nonlinear Distortion. Part I. Test Signals & Measure of Quantification // Proceedings of Telecommunication Universities. 2018. Vol. 4. Iss. 1. PP. 75–84.

Abstract: *The procedure for extracting the signal of nonlinear distortions arising in the sound paths of various devices with the simultaneous presence of frequency and dynamic distortions in the reproduced signal is described. The criteria for estimating their auditory perception and weighting coefficients that take into account the frequency and time response of the signal of nonlinear distortions are proposed. Preliminary experimental data confirming the dependence of the auditory perception of nonlinear distortions as a function of the tone structure of the signal and the spectral distribution of the "noise-to-mask ratio" are obtained.*

Keywords: *multitone signal, nonlinear distortions, simultaneous masking, noise-to-mask ratio, linear multiregression model.*