

Физический и психоакустический анализ цифрового звука с высоким разрешением

Ирина Алдошина



В статье известного российского ученого-акустика Ирины Аркадьевны Алдошиной представлен обзор докладов на конгрессах AES (Audio Engineering society, рус. Международное общество аудиоинженеров), а также другой литературы, посвящённых звуку с высоким разрешением. И.А. Алдошина — профессор, доктор технических наук, заслуженный деятель науки РФ, член Координационного Совета по акустике РАН, председатель Санкт-Петербургского отдела AES, автор более двухсот научных трудов.

Приведены отрывки из статьи. Полный текст — в журнале «Звукорежиссёр» № 4 за 2004 год.

«Суть вопроса заключается в следующем: «Зачем постоянно увеличивать частоту дискретизации в современных системах аудиокоммуникаций (тратя на это огромные средства), если пороги слуховой системы ограничены по частоте диапазоном 20 Гц...20 кГц?»

Попытки ответить на этот вопрос предпринимались многократно, однако однозначного ответа до сих пор нет.

В. Войчик является профессором Университета McGill в Канаде (крупнейшего мирового центра исследований в области психоакустики) и председателем технического комитета AES, поэтому в его распоряжении имеются практически все основные результаты работ, позволяющие в той или иной степени ответить на этот вопрос.

Анализ этих результатов, а также другие соображения, содержащиеся в литературе, и будут представлены в данной статье.

На протяжении почти столетней истории развития технологий звукозаписи главной их целью было нахождение методов сохранения и воспроизведения исполняемой музыки со всеми тончайшими акустическими деталями и максимальной точностью, необходимой для ее восприятия. Термин «натуральность» (который является главной целью создания аппаратуры Hi-Fi) символизирует множество присущих звукозаписи характеристик, которые обеспечивают полное ощущение погружения в музыку с учетом окружающей среды, то есть акустики зала. Такие записи должны обеспечивать восприятие детального взаимодействия между инструментами и исполнителями и реализм в ощущении времени и пространства, то есть чувство нахождения в том акустическом пространстве, где исполняется музыка. Современные цифровые записи с высоким разрешением (высокой частотой дискретизации и большой разрядностью квантования) и стремятся обеспечить слушателей возможностью восприятия большого динамического диапазона, точностью передачи быстрых изменений звука во времени и по другим показателям, то есть точностью передачи деталей пространственно-временнo-спектральной структуры музыки.

Чтобы оценить сложность этой задачи, достаточно представить себе, какие процессы происходят в концертном зале или студии при исполнении оркестром какого-либо музыкального произведения, когда инструменты из различных мест сцены вступают во взаимодействие с акустическими свойствами помещения, находясь при этом во взаимодействии друг с другом. Способность слуховой системы услышать каждый инструмент на его реальном месте, оценить перспективу и влияние помещения с учетом быстрого изменения всех этих взаимодействий во времени зависит, прежде всего, от уникальных возможностей слухового аппарата. Однако при передаче всей пространственной картины ее восприятие зависит и от разрешающей способности записывающих, передающих и воспроизводящих систем. <...>

В многочисленных докладах на конгрессах AES и в дискуссиях на заседаниях технического комитета «Аудио с высокой разрешающей способностью» высказывались мнения, что дальнейший

прогресс в улучшении «прозрачности» и создании «ощущения присутствия» в системах звукозаписи может быть достигнут за счет увеличения разрешающей способности цифровых систем, то есть увеличения частоты дискретизации, и, следовательно, расширения частотного диапазона передаваемого звука за пределы 20 кГц.

Анализ накопленных знаний по данной проблеме позволяет сказать, что этого недостаточно. Учитывая сложность звукового сигнала и свойства слуховой системы, можно утверждать, что только повышение разрешающей способности передающих систем во всех областях (временной, спектральной, пространственной и динамической) может помочь в решении данной проблемы. По крайней мере, **уже сейчас кажется очевидным, что высокая разрешающая способность во временной области является наиболее важной для обеспечения прозрачности звучания.**

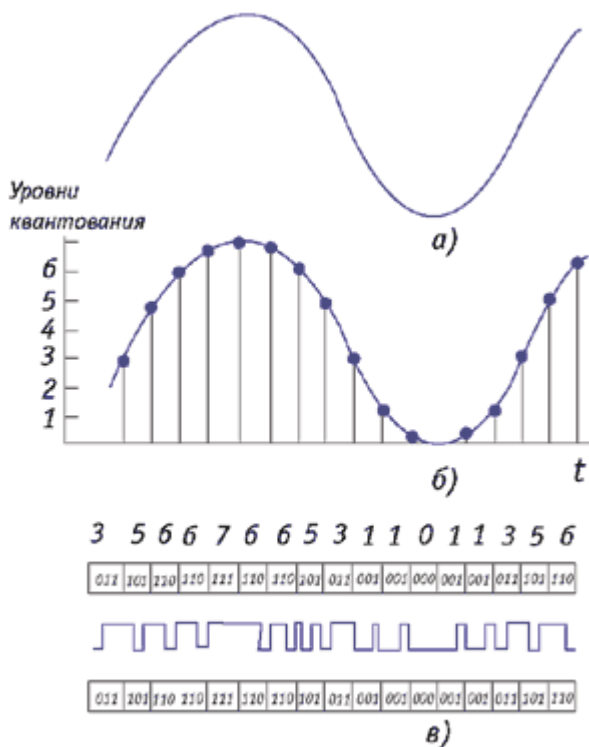


Рис. 1. Преобразование аналогового сигнала в цифровой

Как известно, для превращения аналогового (непрерывного) сигнала в цифровой (дискретный) необходимо выполнить следующие операции: дискретизацию, квантование и кодирование (рисунок 1). Для их выполнения во всех цифровых устройствах (компьютерах, магнитофонах, проигрывателях и пр.) используется аналого-цифровой преобразователь АЦП (ADC), структурная схема которого показана на рисунке 2. В соответствии с теоремой Котельникова (Найквиста) или «теоремой семплирования», для преобразования аналогового сигнала с верхней частотой f_v (Гц) в цифровой без потери информации необходимо, чтобы частота дискретизации, т. е. число отсчетов (семплов в секунду) было не меньше, чем $2 \times f_v$ (Гц). Используемое цифровое слово, число двоичных цифр в котором равно числу выбранных M (бит), представляет мгновенное значение входного сигнала, при этом сигнал квантуется на $2M$ уровней.

Таким образом, теорема семплирования требует, чтобы частота дискретизации была выбрана достаточно высокой $f_d > 2f_v$, при этом сигнал должен оставаться почти постоянным в момент семплирования. Обязательность использования низкочастотного фильтра, который стоит во всех АЦП, не оговаривается, но для предотвращения появления лишних частот в спектре во всех цифровых устройствах стоит антиэлайзинговый фильтр, обрезающий сигнал на частоте $f_d/2$.

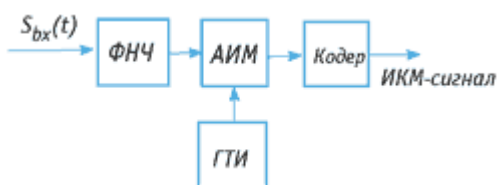


Рис. 2. Структура АЦП

При постоянно меняющейся временной структуре музыкального сигнала временная точность и высокая скорость семплирования, а также точность квантования являются обязательной основой для передачи субъективных признаков звука, в том числе его прозрачности.

В современных цифровых устройствах используются в настоящее время (и будут использоваться в ближайшем будущем) следующие значения основных параметров:

частота дискретизации	верхняя граница частотного диапазона	временной интервал между семплами
44,1 кГц	19 кГц	22,7 мкс
48 кГц	20 кГц	20,8 мкс
96 кГц	35 кГц	10,4 мкс
192 кГц	75 кГц	5,2 мкс
384 кГц	150 кГц	2,6 мкс

Для формата DSD-2 (лабораторные испытания) временной интервал между семплами составляет 0,17 мкс.

Одним из основных направлений исследований в современной психоакустике и является вопрос: «Достаточна ли такая скорость отсчетов или нет, и каков необходимый предел ее увеличения?»

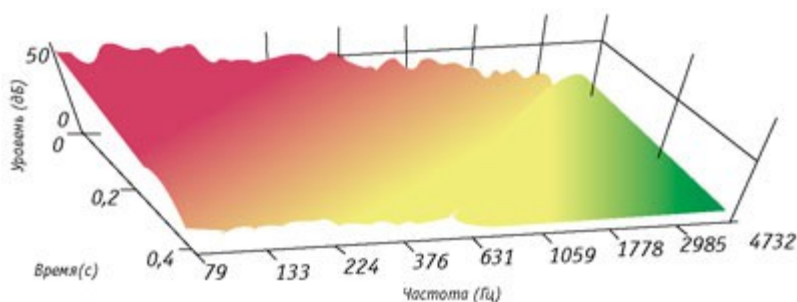


Рис. 3. Запись короткого удара барабана конденсаторным микрофоном

Запись сигнала в любой системе начинается с микрофона (рисунок 3), представляющего собой полосовой фильтр, который уже сам обладает определенными фазовыми и переходными искажениями, приводящими к дисперсии и размытию сигнала во временной области. Данные об этих искажениях редко приводятся в каталогах на микрофоны, однако большой комплекс исследований, выполненный за последние годы, позволил установить значительную разницу по этим параметрам между динамическими и конденсаторными микрофонами. Для конденсаторных микрофонов получены значения **атаки в несколько микросекунд**, в то же время спад переходных процессов достигает нескольких сотен микросекунд. Была **доказана также важность фазовой линейности микрофонов не только внутри, но и за пределами звукового диапазона (в ультразвуковой области)** для обеспечения минимального размытия переходных характеристик музыкального входного сигнала.

Затем аналоговый сигнал, подвергающийся преобразованию в цифровой, обрабатывается низкочастотным фильтром на входе АЦП (антиэлайзинговый фильтр). Этот фильтр также служит причиной дисперсии импульсных характеристик входного сигнала за счет неравномерности АЧХ и ФЧХ

в полосе пропускания, крутизны кривых спада в переходной полосе и фазовой нелинейности (рисунок 4).

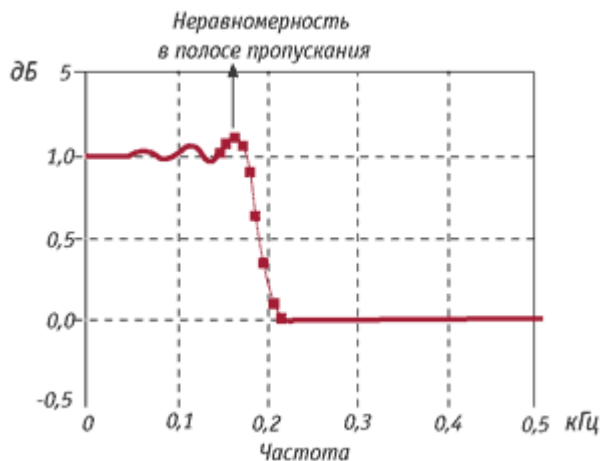


Рис. 4. Частотная характеристика ФНЧ

Такие искажения приводят к временной дисперсии входного сигнала и означают, что каждый мгновенный семпл на выходе будет содержать элементы информации от предыдущих семплов (количество которых зависит от характеристик фильтра). Так как музыкальный сигнал представляет быстро изменяющийся во времени поток с резкими, короткими импульсами, то такая дисперсия и размывание оказывают определенный эффект на слуховое восприятие, особенно для опытного и внимательного слушателя с хорошим музыкальным слухом.

Акустические музыкальные сигналы обладают сверхбыстрой нестационарной временной и динамической структурой, которая обуславливается различными причинами, в частности, быстрой атакой реальных музыкальных инструментов, наличием большого количества ультразвуковых составляющих в спектре многих инструментов, возникновением коротких реверберационных временных задержек в помещении и др. (рисунок 5).

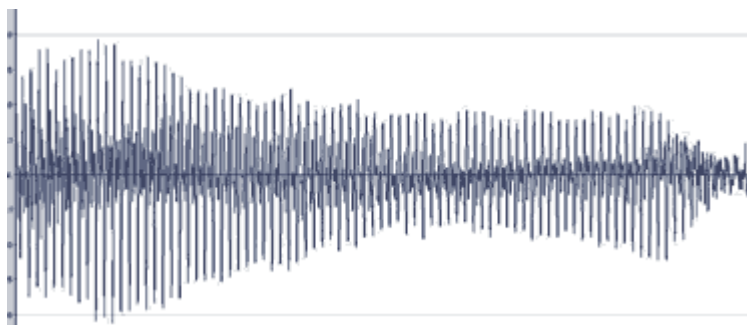


Рис. 5. Временная структура музыкального сигнала (скрипка)

Например, при исполнении форте на трубе звук может достичь пиков 120...130 дБ за 10 мкс, на цимбалах за 7...10 мкс можно получить резкий подъем до 136 дБ. Отсюда следует, что временной интервал семплирования, используемый при записи CD, равный 22,7 мкс явно недостаточен. **Чтобы отследить такие быстрые изменения в сигнале, интервал должен быть не больше 1 мкс.**

Измерения, выполненные с помощью современной прецизионной аппаратуры (специальные измерительные микрофоны V&K 4135, АЦП с частотой 195 кГц и др.), позволили выявить в спектрах музыкальных инструментов большое количество ультразвуковых составляющих. Например, в спектре трубы (нота четвертой октавы 4466 Гц) отчетливо регистрировались составляющие до 40 кГц с уровнем до 60 дБ, в спектре скрипки до 100 кГц, цимбал до 60 кГц с уровнем до 90 дБ. Наличие таких высокочастотных составляющих влияет на временную структуру сигнала и может оказывать косвенное влияние на прослушивание.

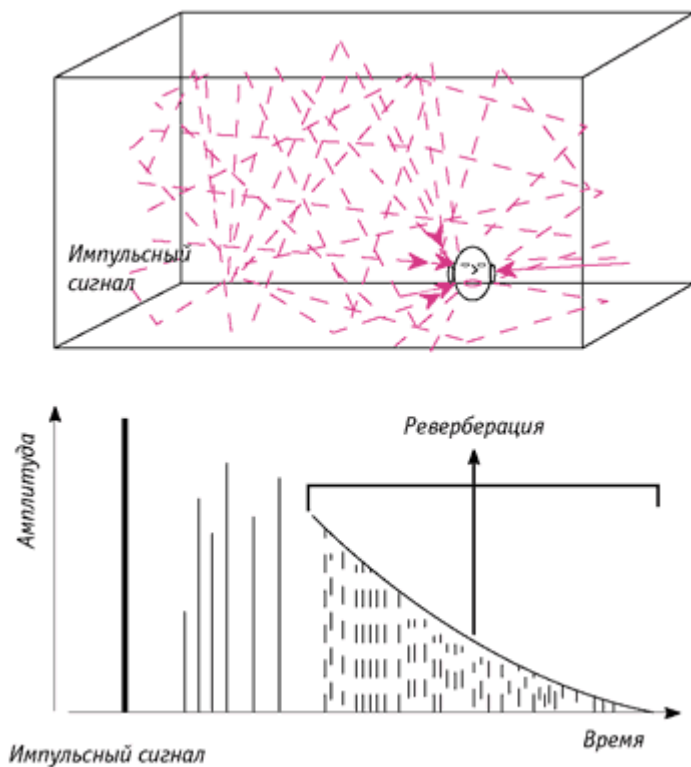


Рис. 6. Реверберационный процесс в помещении

Запись реального реверберационного процесса без потери информации также представляет огромные трудности. Когда источник звука излучает сложный нестационарный музыкальный сигнал, каждый микрофон, установленный в различных точках помещения, «схватывает» сложный отраженный сигнал. Причем дополнительно прибывающие сигналы, измененные по амплитуде и фазе за счет отражений от различных поверхностей, приводят к экспоненциальному возрастанию общего уровня энергии, поступающей на микрофон. При выключении сигнала происходит спад общего уровня, который обычно характеризуется временем реверберации (временем, за который произошел спад уровня сигнала на 60 дБ). Если выполнить детальный анализ процесса спада при помощи импульсного сигнала (рисунок 6), то можно отметить, что первые отраженные сигналы приходят с большой амплитудой и большим разрешением по времени и могут быть хорошо конвертированы в цифровой сигнал. Однако поздние отражения создают огромные проблемы, так как в короткий отрезок времени происходят очень быстрые динамические изменения сигнала.

Для помещения объемом 1000 м³ число отражений через одну секунду после начала реверберационного процесса будет составлять 511,287 отр/с. Это значит, что отраженные лучи будут прибывать с интервалом меньше 2 мкс, вызывая соответствующие флуктуации в выходном сигнале микрофона. Естественно, при временном интервале отсчета семплов 22,7 мс, как для компакт-диска, эти флуктуации никак не смогут быть зарегистрированы.

Тщательные измерения показали также, что в этих отраженных сигналах происходят быстрые амплитудные и фазовые сдвиги и быстрые нерегулярные изменения частоты (частотный джиттер). Когда два или несколько микрофонов распределены в помещении, и при этом еще происходит многодорожечная запись, то эти сложные временные соотношения между сигналами с частотной модуляцией при бинауральном прослушивании создают сдвиги междуушной временной разницы. Как оказалось, слух к этим бинауральным сдвигам (называемым «бинауральный джиттер») очень чувствителен, даже если они составляют доли микросекунд! Кроме того, даже легкое движение исполнителей, воздуха, слушателей и пр. создают дополнительные изменения временных признаков в звуковом сигнале, которые могут восприниматься слуховой системой. <...>

Было показано, что слух особенно чувствителен к флуктуациям временных междуушных задержек в процессе спада сигнала (более чувствителен, чем во время атаки), что особенно важно для слуховой оценки реверберационных процессов в помещении. Минимальная тестируемая разница в двух ушах составляет 6 мкс.

Интересно, отметить, что раньше считалось, будто чувствительность слуха к временным сдвигам сигналов сохраняется, в основном, в диапазоне до 1500 Гц, однако последние результаты показали, что чувствительность к временным различиям сохраняется вплоть до высоких частот для амплитудно-модулированных сигналов, при этом слух извлекает дополнительную информацию из анализа флуктуаций огибающей во времени.

Анализ именно временной разности (а не интенсивностной) является определяющим при локализации сложных сигналов, частично маскирующих друг друга, что типично при восприятии музыкальных сигналов от множественных источников.

Исследования по оценке слышимости упомянутого выше «бинаурального джиттера» (случайных частотных модуляций) показал, что, хотя чистые тоны выше 1400...1600 Гц не могут быть латерализованы, с добавлением случайных частотных модуляций можно выполнить латерализацию сигналов даже при междуушной временной разнице всего в 1,5 мкс. Улучшение бинауральной латерализации к высоким частотам при добавлении джиттера позволяет предположить, что слух отслеживает не только бинауральные временные различия между огибающими сигналов, но и временную междуушную разницу в каждом временном цикле. Удивительно, но было установлено, что слуховая система замечает разницу между чистым импульсным сигналом и сигналом с девиацией всего в 0,2 мкс! Такая высокая бинауральная чувствительность к джиттеру объясняет, почему тонкие частотные флуктуации во время реверберационного процесса в помещении могут быть слышимы. Неизвестно, однако, при какой скорости и уровне флуктуации задержки отраженных сигналов перестают быть слышимыми.

Очевидно, что усредненная импульсная характеристика помещения (дискретизованная и кодированная в цифровой сигнал) не учитывает тонкую структуру и уникальность этих флуктуаций и «затемняет» ощущение присутствия в акустическом окружении. Для этого требуется высокое временное разрешение, частота и точность семплирования, чтобы сохранить каждый бит в описании неоднородностей импульсных характеристик. Неудивительно, что воспроизведение реверберационных процессов представляло значительные трудности для первых цифровых систем из-за потери пространственности (пространство сворачивается в точку) за счет недостатка цифрового разрешения для передачи тончайших временных нюансов в реверберационном процессе. Очевидно, что разрешение в 22,7 мкс (при 44,1 кГц), используемого в компакт-дисках, для этого также явно недостаточно. <...>

Слуховая система способна аккуратно обрабатывать информацию о звуках с задержками до 33 мс, следовательно, первые дискретные отражения воспринимаются и обрабатываются слуховой системой с большой точностью.

Джиттер и модуляционные искажения имеют место во всех звеньях звукозаписи (микрофонах, магнитофонах, громкоговорителях). Наличие джиттера приводит к перераспределению энергии в боковые полосы от основной частоты и может восприниматься на слух как появление некоторой шумовой окраски чистого тона.

Джиттер возникает в аналоговых магнитофонах из-за непостоянной скорости магнитной ленты (в диапазоне от 0,2 Гц до 200 Гц). Кроме того, за счет трения в лентопротяжном механизме возникают продольные колебания ленты (флаттер), что приводит к множественному появлению боковых полос в высокочастотной части спектра. Это обогащает звуковой сигнал некогерентным шумом и создает особое «аналоговое» качество звука (ценимое многими профессионалами). Появление этого флаттер-шума приводит к временным сдвигам сигнала (временным ошибкам), которые оказались равными 10,4 мкс. Отсюда следует, что при переносе на компакт-диск эти флуктуации будут потеряны, так как семплирование сигнала будет происходить только с интервалом 22,7 мкс. <...>

Слух обладает очень высокой разрешающей способностью по времени для периферической зоны слуха (как и для зрения). Периферической считается зона за пределами $\pm 30^\circ$ от срединной плоскости. Для фронтального прослушивания сдвиг источника звука от 0 до 15° приводит к появлению междуушной разницы по времени более 200 мс. Сдвиг на те же 15° в боковой зоне (от 90° до 75°) приводит к разнице только в 20 мс. Чувствительность к временным различиям в периферической зоне оказывается выше, чем во фронтальной, что принципиально важно для передачи ощущения окружения звуковой средой. Частота дискретизации, используемая в компакт-дисках с разрешением 22,7 мс, оказывается недостаточной, чтобы «схватить» эти тонкие временные сдвиги, что приводит к неточной локализации в периферической зоне и потере ощущения окружения (и, возможно, ощущения глубины).

За последние годы было проведено много исследований в области анализа слышимости ультразвуковых частот. Интересные результаты получены в Японии (проф. Ямомото) где было установлено, что подмешивание ультразвуковых компонент к сигналам звукового диапазона улучшает их слуховое восприятие, в то же время подача одних ультразвуковых компонент не создает слухового ощущения. Несколько авторов описывают эксперименты, подтверждающие способность к восприятию ультразвуковых частот за счет костной проводимости (утверждается даже, что люди с потерей слуха могут за счет этого воспринимать ультразвуковую речь, что кажется уж совсем невероятным).

Примечание. Также как и обрезание ультразвуковых компонент сигнала не способствует улучшению качества звука (кроме улучшения липовых показателей по THD).

Эффективность слуховой системы существенно увеличивается с опытом. Например, у опытных звукорежиссеров она существенно выше, чем у обычных слушателей. <...>

Таким образом, анализ показывает, что музыка, исполняемая в помещении, создает сложный звуковой сигнал, который соответствует чрезвычайно тонким и сложным способностям слуховой системы к его анализу. Процесс записи звука должен иметь разрешающую способность, соответствующую как сигналу, так и возможностям слуховой системы. Только когда с помощью новых технологий с высокой разрешающей способностью будет достигнута полная «прозрачность» звука, слушатель сможет ощутить полное погружение в реальную звуковую атмосферу. Технические возможности и эстетические принципы звукорежиссеров могут усилить это чувство погружения за счет расширения перспективы, приближения части звуковой панорамы, вызывающей наибольшее внимание, и других приемов. **Чтобы достичь этого контроля над «эффектом присутствия», требуется высокая временная разрешающая способность всех систем записи и обработки звука, соответствующая уникальным возможностям бинауральной слуховой системы.**

Прогресс в цифровой технике за последние 20 лет характеризуется повышением прозрачности звука и снижением воспринимаемых искажений.

Низкая разрешающая способность первых цифровых систем приводила к характерным искажениям, например, при записи фортепиано — ненатуральная атака и смазывание среднечастотных звуков, при записи ударных — чрезмерно подчеркнутые переходные, нарушение пропорций. При записи оркестра проявлялся недостаток ясности и четкости каждого компонента партитуры. При записи в реверберирующем помещении исчезали тонкие нюансы процесса реверберации и ощущения пространственности — звуки скорее располагались в плоскости, а не распределялись по глубине, происходило восприятие громкости звука без ощущения его объема...

Новые достижения в сверхбыстрой скорости семплирования (то есть увеличении частоты дискретизации) дали цифровому звуку возможности временного и частотного разрешения, близкого к возможностям слуховой системы, а переход на многобитное кодирование 48 бит и более позволил получить высокое динамическое разрешение. Технология «синтеза волнового поля» (Wave Field Synthesis), когда звук исходит от очень большого количества излучающих поверхностей, окружающих слушателя, создает возможности для высокого пространственного разрешения.

Должен быть достигнут, по-видимому, какой-то оптимальный баланс между этими категориями (временной, частотной, динамической, пространственной и др.), при котором увеличение разрешения в одной из них может позволить уменьшить его в другой. Однако **временное разрешение кажется определяющим, поскольку сверхвысокая частота дискретизации повышает слуховую оценку и в других областях.**

Улучшение прозрачности требует, чтобы звукозаписывающие и звуковоспроизводящие системы не создавали модуляционных искажений, которые могут маскировать сложные модуляционные процессы, присутствующие в музыке. Уникальные характеристики джиттера (частотных флуктуаций) в записываемом звуке также не должны затемняться общими модуляционными искажениями записывающих систем, чтобы не нанести ущерба отчетливо различающимся модуляциям индивидуальных звуков.

Таким образом, высокая разрешающая способность звукопередающих систем во временной, спектральной, пространственной и динамической областях совместно определяют качество воспринимаемой музыки и речи, при этом разрешение во временной области имеет доминирующее значение для слухового восприятия.

Все представленные выше результаты не дают, конечно, окончательного ответа на поставленный в начале статьи вопрос, но позволяют еще раз подчеркнуть: слуховая система представляет собой сложный преобразователь, обладающий нелинейными свойствами как на высоких, так и на низких уровнях сигнала, и поэтому однозначного соответствия результатов по слуховым порогам в частотной и временной области не существует (поскольку только в идеальных линейных системах может быть адекватный переход, например, с помощью преобразования Фурье, из одной области в другую).

Поэтому анализ чувствительности слуховой системы к тонким изменениям структуры сигнала должен выполняться для различных категорий: временной, частотной, динамической и т. д. Соответственно, усовершенствование способности цифровой аппаратуры к тонкой передаче характеристик только в одной области (например, в частотной) недостаточно — необходимо комплексное улучшение разрешающей способности технических систем по всем направлениям, что и пытается реализовать современная аудиотехника.»