ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ

Федеральное государственное образовательное бюджетное учреждение высшего профессионального образования «САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ им. проф. М. А. БОНЧ-БРУЕВИЧА»

Ю. А. Ковалгин

ПСИХОАКУСТИКА И КОМПРЕССИЯ ЦИФРОВЫХ АУДИОДАННЫХ

СПбГУТ)))

САНКТ-ПЕТЕРБУРГ
2012

УДК 621.396.97
ББК 32.884.8
К 56

Рецензенты:
доктор технических наук, профессор кафедры акустики и звукотехники Санкт-Петербургского государственного университета кино и телевидения *Ш. Я. Вахитов*,
доктор технических наук, профессор, заведующий кафедрой электротехники и технической электроники Санкт-Петербургского государственного университета кино и телевидения *А. В. Кривошейкин*

Рекомендовано к печати научно-технической комиссией ученого совета СПбГУТ

Ковалгин, Ю. А.
К 56 Психоакустика и компрессия цифровых аудиоданных: [монография] / *Ю. А. Ковалгин*. – СПб. : Издательство СПбГУТ, 2012. – 300 с.
ISBN 978-5-89160-080-5

Рассмотрены свойства слуха, лежащие в основе компрессии цифровых аудиоданных, новейшие алгоритмы обработки звуковых сигналов в системах цифрового телерадиовещания, кинематографа, шоу-бизнеса, такие как *MPEG*-4 *ISO/IEC 14496*-3, *SBR*, *Parametric Stereo*, *MPEG D Surround*, *Dolby AC*-3,*apt-X*100, *ATRAC*, а также качество алгоритмов компрессии.

Для специалистов, интересующихся современными цифровыми технологиями телерадиовещания и аудиотехники, а также для аспирантов, магистров и студентов университетов, обучающихся по направлениям 210700 «Инфокоммуникационные технологии и системы связи» и 210400 «Радиотехника».

Kovalgin Yu.A.
Psychoacoustics and compression (encoding) of digital audio-data: [monograph] / Yu. A. Kovalgin. – SPb.: SPbSUT, 2012. – 300 p.

The book considers the properties of human hearing, forming the basics of compression (encoding) of digital audio-data, and the newest algorithms to process the audio-signals in digital-broadcasting television (TV) systems, in cinemas and for show-business, such as *MPEG*-4 *ISO/IEC 14496*-3, *SBR*, *Parametric Stereo*, *MPEG D Surround*, *Dolby AC*-3,*apt-X*100, *ATRAC*, as well as the quality of these compression (encoding) algorithms.

The book is intended for all specialists interested in exploring the modern digital technologies of TV-broadcasting and the audio-techniques, as well as for Ph.D. students, university master and/or bachelor students being graduated in “Information and communication technologies and systems” (210700) and/orin “Radio-technique” (210400).

УДК 621.396.97
ББК 32.884.8

ISBN 978-5-89160-080-5

© Ковалгин Ю. А., 2012

© Федеральное государственное образовательное бюджетное учреждение высшего профессионального образования «Санкт-Петербургский государственный университет телекоммуникаций» им. проф. М. А. Бонч-Бруевича», 2012

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

При *первичном* кодировании в студийном тракте используется, как правило, равномерное квантование отсчетов звукового сигнала (ЗС) с разрешением ΔA = 16...24 бит/отсчет при частоте дискретизации $f_Д$ = 44,1...96 кГц. В каналах студийного качества обычно ΔA = 16 бит/отсчет, $f_Д$ = 48 кГц, полоса частот кодируемого звукового сигнала ΔF=20...20000 Гц. Динамический диапазон такого цифрового канала составляет около 54 дБ. Если $f_Д$=48 кГц и ΔA = 16 бит/отсчет, то скорость цифрового потока при передаче одного такого сигнала равна v = 48x16 = 768 кбит/с. Это требует суммарной пропускной способности канала связи при передаче звукового сигнала форматов 5.1 (*Dolby Digital*) или 3/2 плюс канал сверхнизких частот (*Dolby Surround*, *Dolby-Pro-Logic, Dolby THX*) более 3,840 Мбит/с. Но человек со своими органами чувств способен сознательно обрабатывать существенно меньшую часть передаваемой информации. Поэтому можно говорить о присущей первичным цифровым звуковым сигналам значительной избыточности.

Различают *статистическую* и *психоакустическую избыточность* первичных цифровых сигналов.

Статистическая избыточность обусловлена наличием корреляционной связи между соседними отсчетами временной функции звукового сигнала (ЗС) при его дискретизации. Для ее уменьшения применяют достаточно сложные алгоритмы обработки. При их использовании потери информации нет, однако исходный сигнал оказывается представленным в более компактной форме, что требует меньшего количества бит при его кодировании. Важно, чтобы все эти алгоритмы позволяли бы при обратном преобразовании восстанавливать исходные сигналы без искажений. Это компрессия цифровых аудиоданных без потерь. Наиболее часто для этой цели используют ортогональные преобразования. Оптимальным с этой точки зрения является преобразование Карунена—Лоэвэ (ПКЛ). Оно обеспечивает представление коэффициентов преобразования в виде последовательности с некоррелированными элементами. Но его реализация при кодировании реальных ЗС вызывает серьезные затруднения, ибо отсутствует быстрый алгоритм вычисления коэффициентов ПКЛ. Расчет коэффициентов ПКЛ требует значительных вычислительных затрат. Поэтому на практике обычно используют субоптимальные преобразования, для которых разработаны быстрые вычислительные алгоритмы. Коэффициенты этих преобразований декоррелированы между собой не полностью. Что касается звуковых сигналов, то чаще всего для этой цели используется *модифицированное дискретное косинусное преобразование* (МДКП), незначительно уступающее ПКЛ по эффективности. Важно также, что для реализации МДКП разработаны быстрые вычислительные алгоритмы. Кроме того, между коэффициентами преобразования Фурье и коэффициентами МДКП существует про-

стая связь, что позволяет представлять результаты вычислений в форме, достаточно хорошо согласующейся с работой механизмов слуха.

Появляется все большее число работ, посвященное вычислению и последующему кодированию коэффициентов ПКЛ. Растет также и число работ, связанных с применением *вейвлет-преобразований*, хотя в большей части публикаций подвергается сомнению его эффективность применительно к кодированию высококачественных ЗС.

Уменьшить скорость цифрового потока позволяют методы кодирования, учитывающие статистику звуковых сигналов (например, вероятности появления уровней ЗС разной величины). Примером такого подхода являются *коды Хаффмана*, где наиболее вероятным значениям сигнала приписываются более короткие кодовые слова, а значения отсчетов, вероятность появления которых мала, кодируются кодовыми словами большей длины.

Именно в силу этих двух причин в наиболее эффективных алгоритмах компрессии цифровых аудиоданных кодированию подвергаются не сами отсчеты ЗС, а коэффициенты МДКП и для их кодирования используются кодовые таблицы Хаффмана. Заметим, что число таких таблиц достаточно велико и каждая из них адаптирована к звуковому сигналу определенного жанра. Также достаточно часто при кодировании используют процедуру группирования. В этом случае при кодировании квантованные отсчеты ЗС или коэффициенты МДКП объединяют в группы. При этом каждую такую группу кодируют одним кодовым словом. Конечно, длина этого слова оказывается большей, чем при кодировании отдельных элементов группы. Однако среднее число бит, приходящееся на кодирование каждого элемента группы, оказывается меньшим, чем при независимом кодировании входящих в нее элементов.

Даже при использовании достаточно сложных процедур обработки устранение статистической избыточности ЗС позволяет уменьшить требуемую пропускную способность канала связи лишь на 15–25%, в редком случае на 30% по сравнению с ее исходной величиной, что никак нельзя считать значительным достижением.

После устранения статистической избыточности скорость цифрового потока при передаче ЗС и возможности человека по их обработке существенно отличаются. Это свидетельствует также о наличии психоакустической избыточности первичных ЗС и, следовательно, о возможности ее уменьшения. Наиболее перспективными с этой точки зрения оказались так называемые *перцепционные методы*, учитывающие такие свойства слуха, как *маскировка, предмаскировка* и *послемаскировка.* Если известно, какие доли (части) звукового сигнала ухо воспринимает, а какие нет вследствие маскировки, то можно вычленить и затем передать по каналу связи лишь те части сигнала, которые ухо способно воспринять, а неслышимые доли (составляющие исходного сигнала) можно отбросить (не передавать по каналу связи). Кроме того, сигналы можно квантовать с возможно меньшим раз-

решением по уровню так, чтобы искажения квантования, изменяясь по величине с изменением уровня самого сигнала, еще оставались бы неслышимыми, то есть маскировались бы сигналом.

После устранения психоакустической избыточности точное восстановление формы временной функции ЗС при декодировании оказывается уже невозможным. Это кодирование сигнала с потерями.

Учет закономерностей слухового восприятия ЗС выполняется в блоке психоакустического анализа. Здесь по специальной процедуре для каждого субполосного сигнала рассчитывается максимально допустимый уровень искажений квантования, при котором они еще маскируются полезным сигналом данной субполосы. Блок динамического распределения бит в соответствии с требованиями психоакустической модели для каждой субполосы кодирования выделяет такое минимально возможное их количество, при котором уровень искажений, вызванных квантованием, не превышал бы порога их слышимости, рассчитанного психоакустической моделью. В современных алгоритмах компрессии используются также специальные процедуры в виде итерационных циклов, позволяющие управлять величиной энергии и формой спектра искажений квантования в субполосах кодирования при недостаточном числе доступных бит. Эта ситуация возникает при малой установленной скорости цифрового потока, когда число доступных для кодирования выборки бит явно недостаточно.

Работы по компрессии цифровых аудиоданных с целью их последующей стандартизации начались в 1988 году, когда была образована международная экспертная группа *MPEG* (*Moving Pictures Experts Group*). Итогом работы этой группы на первом этапе явилось принятие в ноябре 1992 года международного стандарта *MPEG*-1 ISO/IEC 11172-3 (здесь и далее цифра 3 после номера стандарта относится к той его части, где речь идет о кодировании звуковых сигналов). К настоящему времени достаточное распространение получили также и другие стандарты группы *MPEG*, разработанные позже, например такие, как *MPEG*-2 ISO/IEC 13818-3, 13818-7, *MPEG*-4 ISO/IEC 14496-3, *MPEG D Surround*.

Данная книга содержит два раздела. В первом из них рассматриваются психоакустические особенности алгоритмов компрессии цифровых аудиоданных, учет которых особенно важен. Во втором разделе изложены только алгоритмы компрессии с потерями, такие как *MPEG*-4 и *MPEG D Surround*, разработанные группой *MPEG* для цифрового радиовещания и мультимедийных приложений. В отличие от этого в США был разработан стандарт *Dolby AC*-3 (*A*/52) в качестве альтернативны стандартам *MPEG*. Иные алгоритмы компрессии используются в звуковых системах *DTS* (*Digital Theatre Systems*) и *SDDS* (*Sony Dynamic Digital Systems*), применяемых как альтернатива звуковым системам *Dolby Lab* в кинематографе.

Для каждого алгоритма компрессии приводятся результаты тестирования.

1. ПСИХОКУСТИЧЕСКИЕ ОСНОВЫ КОМПРЕССИИ ЦИФРОВЫХ АУДИОДАННЫХ

1.1. Область слышимых звуков

Следует различать абсолютные и относительные пороги слышимости. Они характеризуют чувствительность слухового анализатора к звуковым воздействиям.

Абсолютный порог слышимости. Это минимальное значение звукового давления $p_{АПС}$, которое способно еще воспринять человеческое ухо при отсутствии мешающих звуков. Обычно абсолютный порог слышимости выражают в дБ по отношению к стандартной величине звукового давления $p_0 = 2\times10^{-5}$ Па (рис. 1.1,*а*). По оси ординат здесь отложены уровни звукового

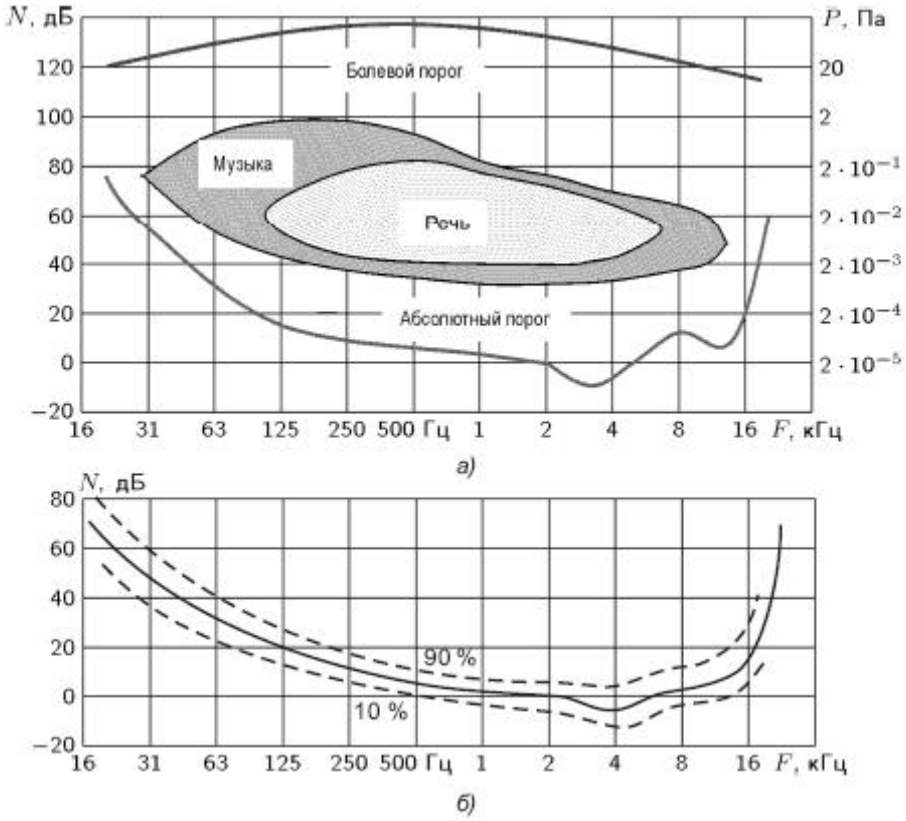


Рис. 1.1. Область слышимых звуков по амплитуде и частоте

сигнала $N_{АПС} = 20\lg(p_{АПС}/p_0)$ в дБ, по оси абсцисс – значения частоты в Гц в логарифмическом масштабе. Пунктирные кривые (рис. 1.1,*б*) дают наглядное представление о разбросе чувствительности слуха для разных людей. Этот разброс минимален на средних частотах. В этой области частот 80% данных всех измерений оказались в пределах диапазона шириной 10 дБ. Полоса разброса экспертопоказаний медленно расширяется в сторону низких частот и заметно быстрее в сторону верхних частот. Видно, что чувст-

вительность слуха к восприятию отдельных спектральных составляющих в сильной степени зависит от частоты (рис. 1.1,*а* нижняя кривая).

Границы воспринимаемого слухом частотного диапазона довольно широки: обычно приводят цифры 20…20000 Гц, реально эти цифры для большинства людей в молодом возрасте равны 30-35….16000-18000 Гц. В области частот от 2 до 5 кГц чувствительность слуха максимальна, она медленно ухудшается в сторону низких частот и заметно быстрее в сторону верхних частот. Те спектральные компоненты полезного сигнала, которые лежат ниже абсолютного порога слышимости на слух не воспринимаются.

Следует заметить, что звуковое давление, возникающее вследствие броуновского движения молекул при температуре 25^0 С, составляет около $5 \cdot 10^{-6}$ Па. Если бы ухо было вдвое чувствительней, оно слышало бы непрерывный шум флуктуаций молекул воздуха и тока крови. Таким образом, чувствительность уха находится на пределе биологической целесообразности.

Пороги слышимости для левого и правого уха даже у вполне здоровых людей различны. Кроме того, результаты измерений зависят от того, что используется при измерениях: громкоговоритель или телефоны. При использовании телефонов абсолютный порог слышимости выше на 5-10 дБ. С возрастом слух людей притупляется и быстрее всего на высоких частотах. Обычно на частоте 10 кГц чувствительность уха у 60 летнего человека на 20 дБ ниже, чему 20- летнего.

Кривая абсолютного порога слышимости от частоты хорошо аппроксимируется уравнением

$$N_{АПС} = 20\lg(p_{АПС} / p_0) = 3{,}64 \cdot F^{-0{,}8} - 6{,}5 \cdot \exp[-0{,}6 \cdot (F - 3{,}3)^2] + 10^{-3} \cdot F^4 ,$$

где: F – частота в кГц; $p_{АПС}$ – звуковое давление, соответствующее абсолютному порогу слышимости.

Величина абсолютного полога слышимости зависит также и от длительности стимула. Представленные на рис. 1.1 кривые справедливы при длительности звукового воздействия не менее 250 мс. При меньших значениях величина абсолютного порога слышимости повышается. Например, уменьшение длительности воздействующего сигнала с 200 до 20 мс сопровождается повышением абсолютного порога на 10 дБ. Очень короткие звуки воспринимаются как щелчок. Необходимо определенное время воздействия звукового стимула, чтобы можно было определить высоту тона. Длина этого минимального отрезка зависит от частоты: при частоте 50 Гц она равна 60 мс, при частоте 1000 Гц – 10 мс.

Слух человека инерционен, он интегрирует энергию воздействующего сигнала за определенный промежуток времени и только после того, когда накопленная энергия превышает некоторый порог, возникает ощущение

присутствия звука. Для точной оценки его уровня громкости необходим временной интервал около 200 мс.

Болевой порог. При звуковом давлении 10 Па (N = 100 дБ) возникает неприятное ощущение, при давлении 60-80 Па (N = 120…130 дБ) – ощущение давления на уши. При давлении 150…200 Па (около 135..140 дБ) появляется болевое ощущение, это так называемый *болевой порог* (рис. 1.1,а, верхняя кривая). Его величина в существенной меньшей степени зависит от частоты.

Слуховая система человека приспособлена к звукам малой и средней интенсивности с уровнем, не превышающим 90…96 дБ. Звуки с уровнем больше 100 дБ при длительном слушании приводят к изменению порогов слышимости, в конечном итоге к потере слуха. Степень повреждения слуха пропорциональна времени воздействия. Иногда порог чувствительности восстанавливается через 16-20 часов. В последние годы наблюдается снижение слуховой чувствительности у молодежи, что связано с образом их музыкальной жизни, точнее говоря со злоупотреблениями, связанными с длительным прослушиванием музыкальных записей на повышенном уровне громкости.

Допустимое время пребывания человека в условиях громких звуков (что важно для звукорежиссеров) регламентируется международными документами. Например, при уровне N = 90 дБ это 8 часов в день, при N = 95 дБ – соответственно 4 часа, при N = 100 дБ – 2 часа, при N=105 дБ – 1 час в день, при N = 110 дБ – 0,5 часа в день, при N = 115 дБ – 0,25 часа в день. При превышении этих норм могут возникнуть необратимые изменения в слуховой системе, приводящие к понижению чувствительности слуха.

Область слышимости. Кривые абсолютного порога слышимости и порога болевого ощущения (рис. 1.1,*а*) ограничивают область слышимых звуков, определяют динамический диапазон слуха: на средних частотах он составляет не менее 120…130 дБ, на низких и высоких частотах уменьшается тем больше, чем ниже или выше частота сигнала. Здесь же (рис. 1.1,*а*, затененные части) для большей наглядности показаны области, где могут находиться уровни и частоты речевых и музыкальных сигналов в естественных условиях. Видно, что спектры речевых сигналов имеют полосу частот от 35 до 7000 Гц, при этом их уровни лежат в интервале от 40 до 84 дБ; для музыкальных сигналов эти значения соответственно равны 31….1500–18000 Гц и 35…96–100 дБ. В обоих случаях динамический диапазон реальных звучаний существенно зависит от частоты.

Спектральные компоненты, уровень которых лежит ниже абсолютного порога слышимости передавать на приемную сторону системы телерадиовещания нет необходимости, ибо они не воспринимаются слухом.

Динамический диапазон слуха. В тишине чувствительность слуха человека повышается, а в присутствии громких звуков – понижается, слух адаптируется к окружающей среде, поэтому динамический диапазон слуха

не такой большой, как о нем говорят – около 70..80 дБ. Сверху он ограничен давлением 100 дБ, *SPL*, а снизу шумом окружающей среды, составляющим около 35…40 дБ, *SPL*, в тихих помещениях. Этот динамический диапазон может сдвигаться вверх и вниз до 20 дБ. Для комфортного восприятия музыки рекомендуется, чтобы максимальное звуковое давление не превышало 104 дБ, *SPL*, в домашних условиях и 112 дБ, *SPL*, в концертных залах.

1.2. Слуховые пороги

Следует различать абсолютные и дифференциальные слуховые пороги.

Абсолютные слуховые пороги определяются минимальными значениями того или иного параметра звукового стимула (звукового давления, частоты, длительности), при котором возникает слуховое ощущение. Ярким примером здесь является кривая абсолютного порога слышимости (рис. 1.1,*а*).

Дифференциальные слуховые пороги определяют способность слуха обнаруживать и оценивать небольшие изменения того или иного параметра звукового стимула (давления, частоты и т. п.).

Под разрешающей способностью слуха понимают минимальные изменения звукового давления и частоты, которые еще могут быть замечены слухом. Разрешающую способность слуха по амплитуде (едва заметному на слух изменению громкости сравниваемых звуков) и частоте (едва заметное изменение высоты тона) иллюстрируют так называемые дифференциальные слуховые пороги (*JND – just noticeable difference*).

Пороги различимости по амплитуде

Эти пороги измерялись с помощью амплитудной модуляции тоном, как для синусоидальных сигналов различной частоты, так и для шумовых сигналов.

Тональные испытательные сигналы. Наибольшая чувствительность слуха была отмечена при частотах модуляции около 4 Гц. Поэтому все дальнейшие исследования большинством авторов проводились именно при этой частоте. Эксперименты сводились к оценке минимального уровня звукового давления, при котором становились заметными колебания громкости испытательного сигнала, обусловленные амплитудной модуляцией.

Прежде всего, рассмотрим зависимость минимально ощущаемых изменений амплитуды тона от уровня его звукового давления при частоте тона 1000 Гц и частоте модуляции 4 Гц (рис. 1.2,а). Видно, что с увеличением уровня испытательного сигнала величина коэффициента минимально

ощущаемой амплитудной модуляции уменьшается, то есть чувствительность слуха к изменению амплитуды сигнала растет. При уровне звукового

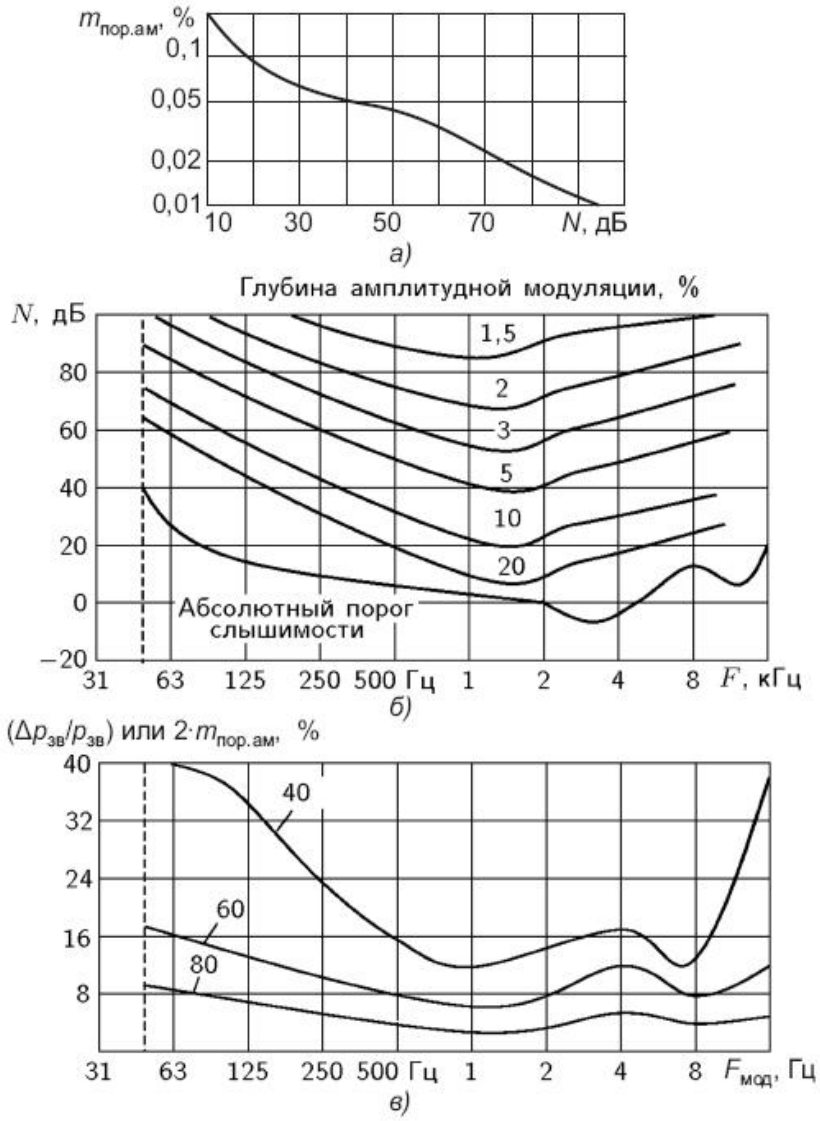


Рис. 1.2. Зависимость коэффициента $m_{пор.ам}$, %, минимально ощущаемой амплитудной модуляции тона частотой 1000 Гц от уровня N, дБ (*а*), кривые равной заметности амплитудной модуляции (*б*) и зависимости относительного изменения минимально ощущаемого звукового давления тона частотой 1000 Гц от частоты модуляции $F_{МОД}$, Гц (*в*)

давления N = 20 дБ коэффициент минимально ощущаемой модуляции составляет m = 10%. При уровне 100 дБ он уменьшается до 1%, то есть возрастает в данном случае в 10 раз. Отсюда следует, что при высоких уровнях звукового давления слух весьма чувствителен к изменению во времени амплитуды тонального сигнала.

Аналогичные данные получены и для модулированных по амплитуде тональных сигналов других частот (рис. 1.2,*б*). Представленные здесь зависимости можно назвать кривыми равной заметности амплитудной модуляции. Здесь по оси ординат отложен уровень сигнала N, дБ, по оси абсцисс –

частота F, Гц, синусоидального сигнала. Параметром каждой кривой является глубина амплитудной модуляции $m_{\text{пор.ам}}$, %. По форме эти зависимости повторяют так называемые кривые равной громкости, о которых будет сказано позже, и кривую абсолютного порога слышимости (нижняя кривая).

Величина порога различимости по амплитуде зависит не только от уровня звукового сигнала, но, как это уже было сказано выше, и от частоты модуляции $F_{\text{мод}}$ (рис. 1.2,*в*), где параметром кривых является уровень звукового давления N, дБ. Эти данные получены при изменении амплитуды тона частотой 1000 Гц. Видно, что при больших уровнях (N = 80 дБ) уже заметно относительное изменение амплитуды звукового давления ($\Delta p_{\text{зв}}/p_{\text{зв}}$, %) на 3%, в то время как при уровне 40 дБ разрешающая способность слуха по амплитуде уменьшается до 10%. Здесь величина $\Delta p_{\text{зв}}$ – полное изменение амплитуды сигнала (в обе стороны), поэтому отношение $\Delta p_{\text{зв}}/p_{\text{зв}}$ в два раза больше соответствующего значения коэффициента амплитудной модуляции m. Заметим также, что с уменьшением уровня тона резче проявляется и частотная зависимость слуха к изменению амплитуды сигнала.

Ход кривых подтверждает, что наш слух наиболее чувствителен к изменению амплитуды сигнала при частоте модуляции 4 Гц.

Оценивая для чистых тонов разрешающую способность слуха по амплитуде, можно сказать, что для области слышимых звуков и для наиболее часто встречаемых уровней от 40 до 80 дБ она составляет 0,5…1 дБ.

Шумовые испытательные сигналы. Намного чаще, чем чистые тоны, в практике встречаются созвучия и шумоподобные сигналы. Созвучия занимаю среднее положение между чистыми тонами и шумами. Глухие согласные, например шипящие звуки, более схожи по форме спектра с фильтрованным шумом. Безусловно, важно знать свойства слуха не только при восприятии чистых тонов, но и шумов, что, в свою очередь, позволит получить сведения о свойствах слуха при восприятии речевых и музыкальных программ.

Широкополосный белый шум. На рис. 1.3,а представлена зависимость коэффициента минимально ощущаемой амплитудной модуляции $m_{\text{пор.ам}}$ белого шума от уровня звукового давления N. В качестве модулирующего сигнала взят тон частотой 4 Гц. Только при уровнях, не превышающих 30 дБ, кривая аналогична рис. 1.2,*а*. При более высоких уровнях $N > 30$ дБ форма кривой уже не зависит от уровня испытательного сигнала. Порог различимости по амплитуде составляет в этой области около 4%. Влияние частоты модулирующего сигнала показано на рис. 1.3,*б*. Видно, что при низких частотах модулирующего сигнала слух в состоянии следить за все ми изменениями уровня. Как и ранее чувствительность слуха максимальна при частоте модуляции 4 Гц. При частотах модуляции превышающих 10 Гц воспринимаемый сигнал становится неровным, клокочущим. Чувствитель-

ность слуха к восприятию амплитудной модуляции ухудшается с повышением частоты модулирующего сигнала $F_{мод}$.

Узкополосный шум. Пусть в качестве модулирующего сигнала взят тон частотой 4 Гц. Если полоса шума достаточно широка, то даже при наличии присущей ему модуляции такой шум воспринимается как равномерный. При этом наличие амплитудной модуляции заметно уже при величине

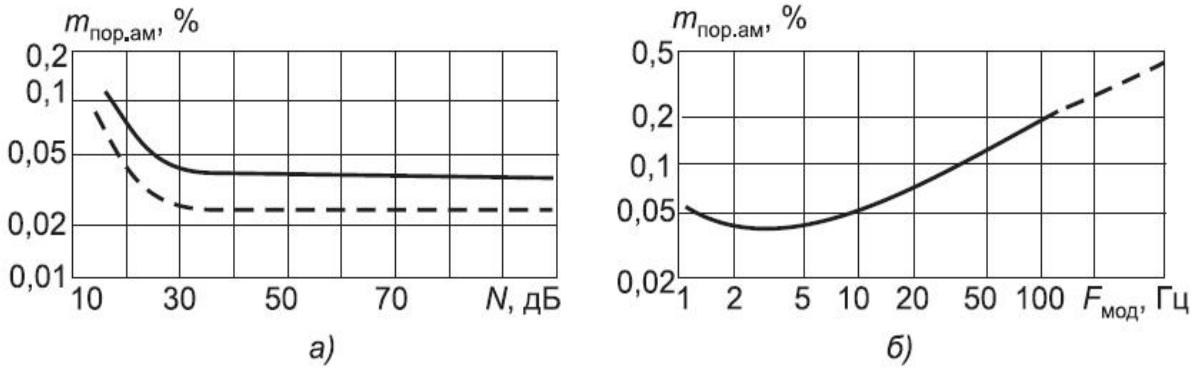


Рис. 1.3.Кривые зависимости минимально ощущаемой амплитудной модуляции белого шума для тонального модулирующего сигнала частотой 4 Гц (*а*, жирная линия – модуляция тоном, штриховая линия – модуляция прямоугольным по форме сигналом) и влияние частоты модулирующего сигнала (*б*)

$m_{пор.ам}$ = 4%. Чем уже полоса частот шума $\Delta F_ш$, тем заметнее на слух становятся его собственные неоднородности, а это затрудняет восприятие самой модуляции. Она становится заметной при большем значении $m_{пор.ам}$, рис. 1.4,*а*. Видно, что при $\Delta F_ш$=50 Гц, модуляция оказывается слышимой, если ее глубина превышает 30%. При полосе белого шума равной 1000 Гц эта величина составляет уже 11 %. При сокращении полосы частот шума до 10 Гц слух обнаруживает его собственную неупорядоченную модуляцию с глубиной 70%, таким же оказывается и пороговое значение глубины амплитудной модуляции, если в качестве модулирующего сигнала взят тон частотой 4 Гц. Данные, представленные на рис. 1.4,*а* справедливы для всех значений средних частот узкополосного шума при условии, что его полоса меньше остается меньше частотной группы (критической полосы слуха). Это термин будет пояснен позже. Если же шум перекрывает несколько критических полос слуха, то амплитудная модуляция распознается при меньших значениях ее глубины. В области частот от 100 до 500 Гц ширина критических полос слуха одинакова, она примерно равна 100 Гц. Величина минимально ощущаемой амплитудной модуляции $m_{пор.ам}$ для шума, полоса частот которого равна критической полосе слуха, составляет 25%. В случае же, когда спектр шума перекрывает пять критических полос слуха, то есть имеет полосу частот 100…500 Гц, величина порога заметности амплитудной модуляции уменьшается до 14%. Заметим, что это значение можно по-

лучить из рис. 1.4,*а*, где оно соответствует ширине полосы шума равной 500 Гц.

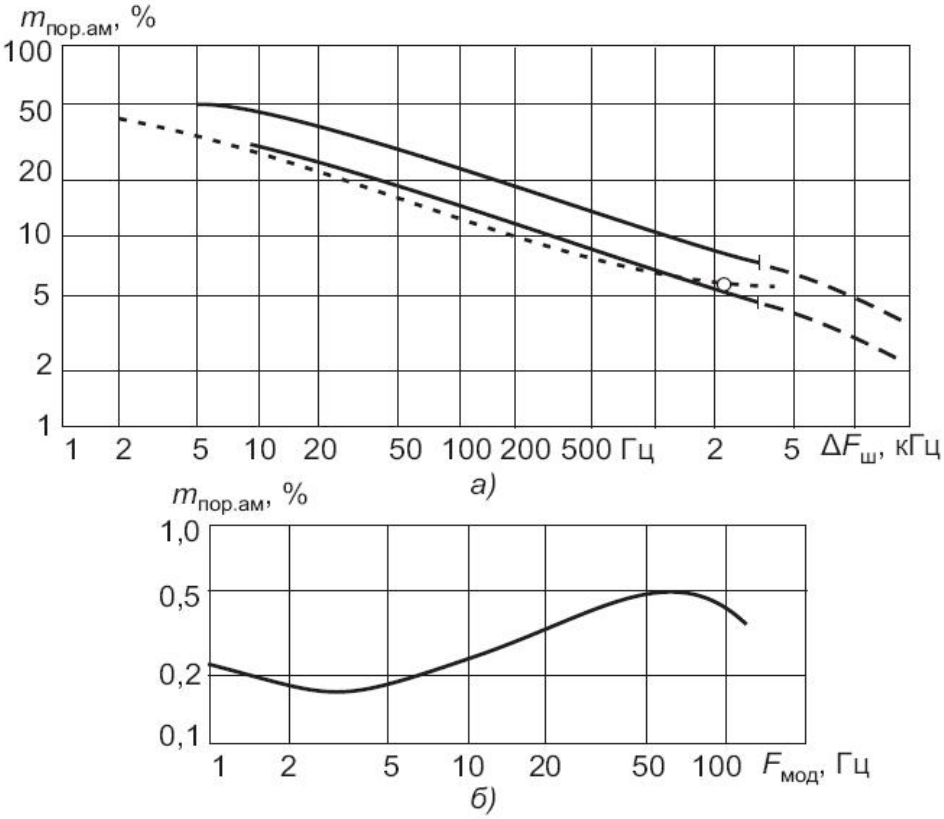


Рис. 1.4. Зависимости минимально ощущаемой амплитудной модуляции узкополосного шума от ширины его полосы при частоте модуляции 4 Гц (*а*) (штриховые участки – полоса шума шире критической, точечная линия – результат вычислений по порогам маскировки) и от частоты модуляции (*б*)

Зависимости минимально ощущаемой (слышимой) амплитудной модуляции от частоты модуляции (рис. 1.4,*б*) имеют такой же характер, как и для белого шума (рис. 1.3,*б*). Отличие состоит лишь в числовых значениях.

Пороги различимости по частоте

При проведении данных экспертиз возможны два подхода. В первом случае слушателю предлагается последовательно во времени сравнить по высоте два тона одинаковой громкости. Меняя частоту одного тона относительно другого, определяют наименьший разнос их по частоте, при котором слушатель еще способен отличить их по высоте. При втором подходе, используя частотную модуляцию синусоидального сигнала тоном, разрешающую способность слуха по частоте оценивают минимальными значениями девиации частоты, еще замечаемой слухом как изменение высоты звука.

Тональные испытательные сигналы. Мгновенные изменения частоты тона воспринимаются на слух как щелчки. Поэтому их исключим из рас-

смотрения и сосредоточим наше внимание на тонах модулированных по частоте. Область повышенной разрешающей способности слуха наблюдается при частотной модуляции синусоидального сигнала тоном частотой около 4 Гц. Так как частота синусоидального сигнала при частотной модуляции тоном изменяется в пределах от $F - \Delta F_д$ до $F + \Delta F_д$, где $\Delta F_д$ – девиация частоты, то градацией раздражения в данном случае будет величина $2\Delta F_д$.

Результаты экспертиз показывают, что величина минимально ощущаемой частотной девиации зависит от частоты модуляции $F_{мод}$ испытательного тона и от его уровня N.

На рис. 1.5,*а* показаны кривые разрешающей способности слуха по частоте. По вертикальной оси отложено минимально ощущаемое относительное изменение частоты тона $\Delta F_п/F$, %; по горизонтальной оси – частота модулируемого синусоидального сигнала F, Гц. Параметром каждой кривой является уровень N, дБ. Частота модулирующего сигнала равна 4 Гц. Из данных кривых следует, что в области низких частот разрешающая способность слуха по частоте определяется абсолютным значением изменения частоты ΔF тона, на высоких частотах – его относительным значением $\Delta F/F$.

На рис. 1.5,*б* представлена зависимость порогового значения девиации частоты тона 1000 Гц от уровня звукового давления N, дБ. Видно, что для малых уровней $N < 30$ дБ пороговое значение девиации частоты, еще замечаемое слухом, существенно возрастает с понижением N. При средних и больших уровнях $N > 30$ дБ эта величина остается практически постоянной.

На рис. 1.5,*в* дана зависимость пороговой девиации частоты $\Delta F_Д$ модулируемого тона от частоты модуляции $F_{мод}$ при уровне звукового давления $N = 70$ дБ. В области частот $F < 500$ Гц пороговая величина девиации практически не зависит от частоты и составляет примерно 1,8 Гц. На частотах выше 500 Гц величина пороговой девиации возрастает почти пропорционально частоте и составляет примерно

$$\Delta F_Д = 3{,}5 \cdot 10^{-3} F \; [(\Delta F_Д/F) = 3{,}5\%].$$

И, наконец, на рис. 1.5,*г* представлено семейство кривых равной заметности частотной модуляции. По оси ординат здесь отложены уровни тональных сигналов N, дБ, по горизонтальной оси их частоты F в Гц, параметром каждой кривой является значение индекса частотной модуляции, %. Частота модулирующего сигнала равна 4 Гц.

В целом, в диапазоне частот до 1 кГц порог различимости по частоте $\Delta F_п$ равен 2…3 Гц (в ряде работ приведена цифра 1 Гц), в полосе частот от 1000 до 10000 Гц это значение равно $(\Delta F_п/F) = 0{,}035$. В литературе, посвя-

щенной измерениям дифференциальных частотных порогов (*JND*) слуха, часто приводят зависимость, представленную на рис. 1.6.

Шумовой испытательный сигнал. Ниже представлены кривые, полученные для белого шума. В шумовом сигнале звуковое давление в каждой

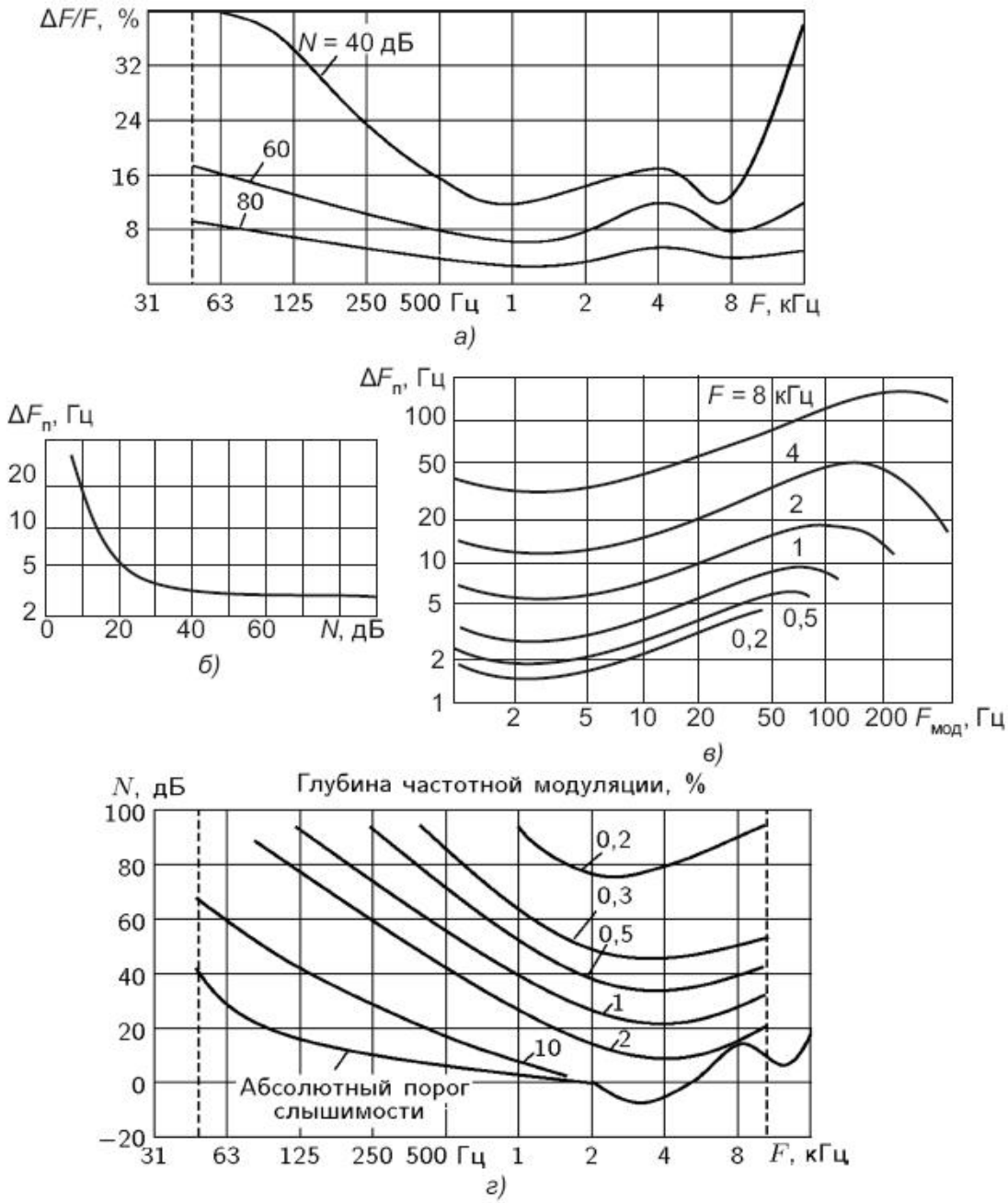


Рис. 1.5. Зависимости минимально ощущаемого относительного изменения частоты $\Delta F_п/F$ от частоты F для трех значений уровня N тона (*а*), минимально ощущаемой девиации частоты $\Delta F_Д$ тона от уровня N (*б*) и от частоты модуляции $F_{мод}$ при постоянном уровне $N = 70$ дБ (*в*) испытательного сигнал, кривые равной заметности частотной модуляции (*г*). Частота модуляции во всех случаях составляет 4 Гц, в случае *б* частота тона составляет 1000 Гц

частотной полосе непрерывно меняется. Поэтому частотная модуляции при таком испытательном сигнале будет слышна лишь в том случае, когда она превосходит эту собственную модуляцию, присущую шуму. Именно по этой причине поровые значения минимально ощущаемой девиации частоты здесь выше, чем для тональных сигналов. В качестве примера на рис. 1.7,*а*

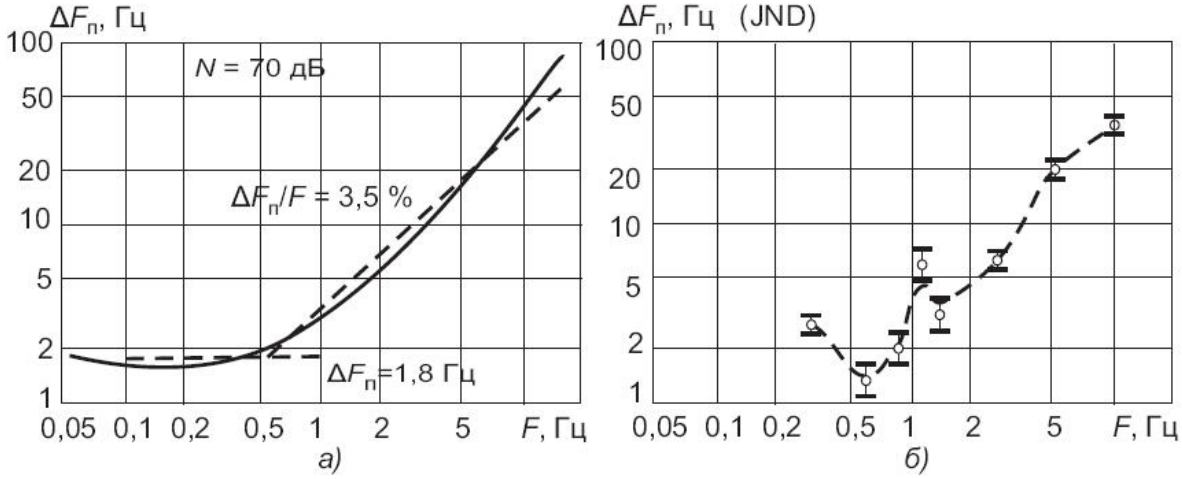


Рис. 1.6. Кривая дифференциального частотного порога в функции от частоты: *а* – по Э. Цвикеру [1.4], *б* – чаще всего используемая в международных документах зависимость

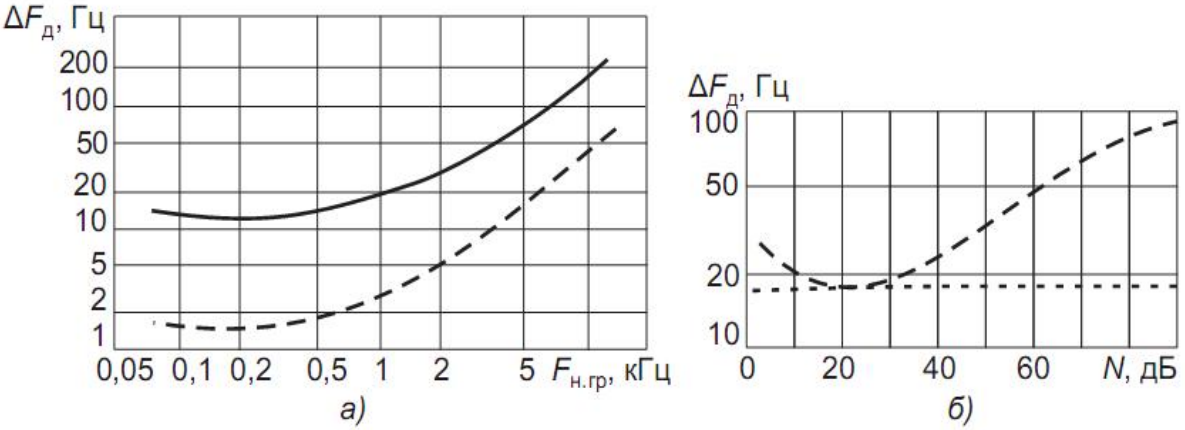


Рис. 1.7.Минимально ощущаемая девиация частоты $\Delta F_Д$ высокочастотного шума от значения нижней граничной частоты $\Delta F_{н.гр}$ (а, сплошная кривая), а также низкочастотного (пунктирная линия) и высокочастотного (точечная линия) шумов с граничной частотой 1 кГц от уровня *N* (*б*). Частота модуляции равна 4 Гц

представлена зависимость минимально ощущаемой девиации частоты $\Delta F_п$ высокочастотного шума в функции от значения его нижней граничной частоты $F_{н.гр}$. Частота модуляции равна 4 Гц. Для сравнения здесь же пунктирной линией показана аналогичная зависимость, полученная тонального сигнала. Видно, что восприятие зависит от частоты: в области частот $F >$ 500 Гц разрешающая способность слуха по частоте уменьшается с ростом частоты, причем это изменение происходит медленнее, чем для чистого тона, что также связано с влиянием собственной модуляции шума. На часто-

тах ниже 500 Гц разрешающая способность слуха по частоте не зависит от значения нижней граничной частоты шума.

Важно, что разрешающая способность слуха по частоте для высокочастотного шума не зависит от его уровня (рис. 1.7,*б*), то есть он ведет себя также как и чистый тон. В отличие от этого для низкочастотного шума величина минимально ощущаемой девиации частоты при возрастании его уровня существенно увеличивается.

Как и ранее, по-прежнему, слух наиболее чувствителен к модуляции с частотами от 2 до 5 Гц.

Разрешающая способность слуха по времени

Слух по своей природе инерционен и для точной оценки того или иного параметра испытательного сигнала (уровня громкости, высоты тона, различия по частоте и т.п.), тех или иных изменений в его временной структуре и в спектре требуется определенное время. Оно оказывается разным при точной оценке тех или иных ощущений. С этой точки зрения также можно говорить о временных дифференциальных порогах слуха.

Наибольшее число исследований, посвященных выявлению данных закономерностей, связано со следующими направлениями:

-оценкой минимального времени, в течение которого слух способен различит два сигнала. Например, для чистых тонов, следующих друг за другом, эта величина составляет около 2 мс, она не сильно зависит от частот сравниваемых тонов и их уровня. Интересно, что если требуется при этом еще и определить, какой из двух сигналов поступает первым, то это время возрастет уже до 20 мс. Для распознавания фонем речи необходимо 35 мс. Для определения высоты тона при низких частотах требуется около 60 мс, при высоких – 15 мс. Для оценки наличия в сигнале нелинейных искажений нужно около 10 мс, для оценки направления на источник звука около 120…150 мс и т.п.;

-оценкой дифференциальной чувствительности слуха к изменению длительности звукового воздействия. В качестве примера рассмотрим зависимость, представленную на рис. 1.8,*а*. Здесь по оси ординат отложено минимально воспринимаемое экспертами различие по частоте $2\Delta F$ двух следующих друг за другом тональных импульсных сигналов одинаковой длительности $t_и$ каждого из них. Параметром кривых является значение частоты F. В ходе исследований эксперту предъявлялись тональные импульсные сигналы с частотами $F_1 = F - \Delta F$ и $F_2 = F + \Delta F$. Кратковременные паузы между импульсами мешающего действия на результаты эксперимента не оказывают. Формы тональных импульсов наиболее пригодные для данной оценки показаны на рис. 1.8,*б*. Она обеспечивает отсутствие щелчков при восприятии.

При уменьшении длительности импульсов минимально ощущаемый частотный интервал возрастает. Так при сокращении длительности импуль-

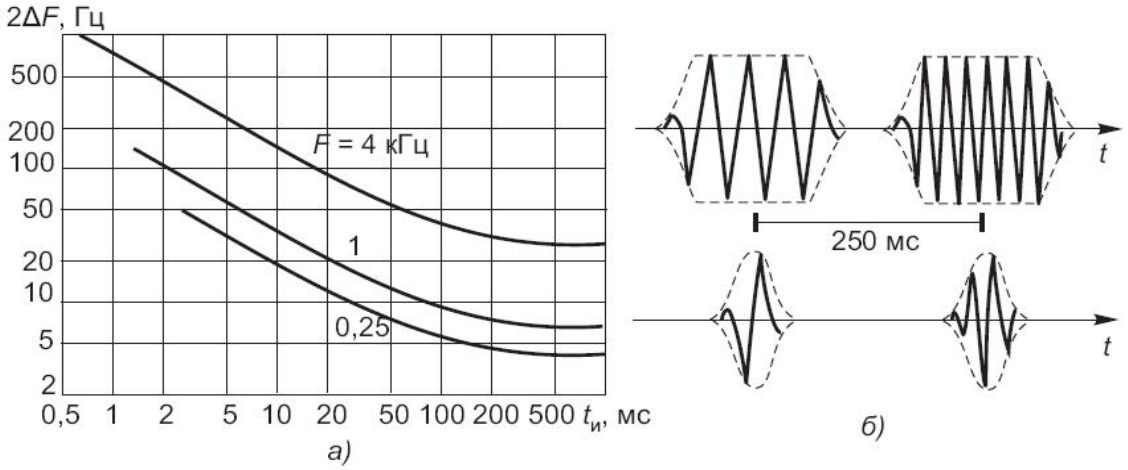


Рис. 1.8.Зависимость минимально ощущаемого частотного интервала (порога различимости по частоте) между тональными импульсными сигналами от длительности импульсов (*а*) и формы тональных импульсов (*б*)

сов с 200 до 5 мс он увеличивается почти в 10 раз для любой из трех частот.

Другой пример представлен на рис. 1.9,*а*. Здесь слушатель должен был сравнить по длительности два сигнала, длительность одного из них составляла Т, а длительность – другого Т + ΔТ. Сигналы предъявлялись экспер-

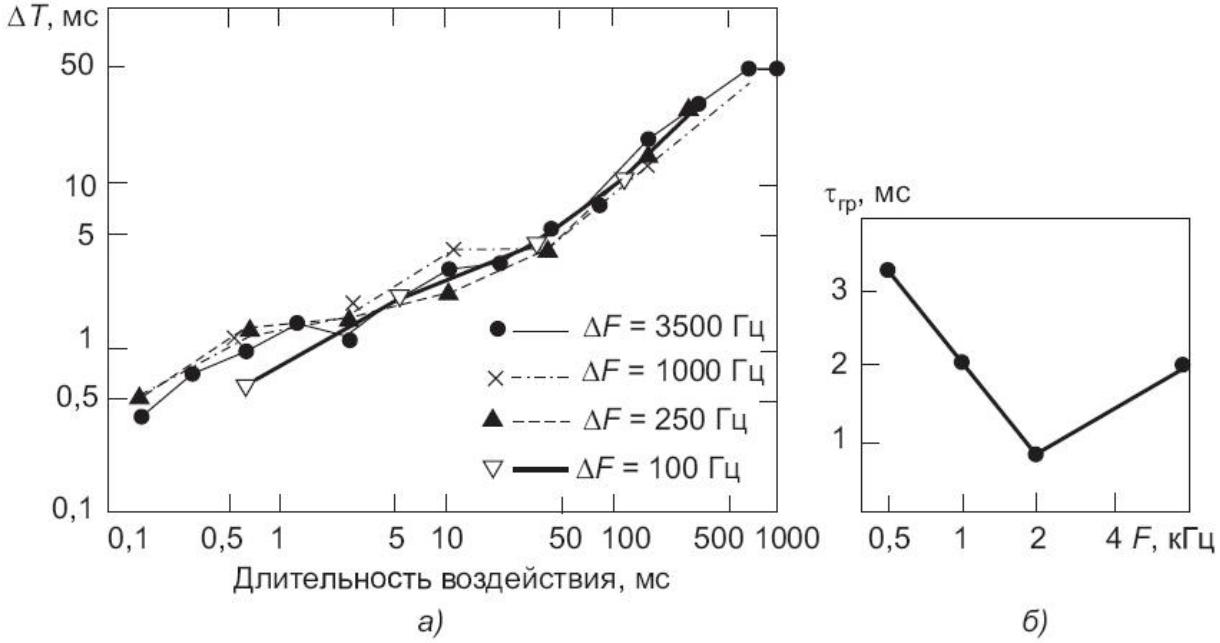


Рис. 1.9.Зависимость дифференциальных слуховых порогов от длительности воздействующих звуков для шумовых сигналов со средней частотой 1000 Гц и с различной полосой частот ΔF (*а*) и пороги слуховой заметности для ГВЗ (*б*)

там в случайном порядке их следования. Эксперты должны были сказать, какой их двух сигналов в паре имеет большую длительность. Величина ΔT менялась ступенями от одной пары к другой также в случайной последовательности. Видно, что пороговое значение ΔT составляет около 50 мс при длительности сигнала T = 960 мс и уменьшается до 0,5 мс при длительности сигнала T = 0,5 мс. При этом отношение $\Delta T/T$ (отношение Вебера) равно 1 при T = 0,5…1,0 мс, равно 0,3 при T = 10 мс и равно 0,1 при T = 50…500 мс. Полученные результаты мало зависят от полосы частот и уровня звука;

-оценкой чувствительности слуха к изменению длительности атаки или спада сигнала звука. Она оказывается равной $\Delta\tau$ = 1 мс для частот ниже 1000 Гц и около 0,5 мс для частот от 1 до 10 кГц. Меньшие значения изменения времени атаки и спада звукового стимула слухом не замечаются;

-оценкой чувствительности к фазовым изменениям в сигнале. Долгое время считалось, что слух практически не чувствителен к изменениям фаз спектральных компонент сложных сигналов. Новейшие исследования показали, что изменение фазовых соотношений сложного сигнала влияет на тембр, четкость, высоту звука. Наиболее заметны эти изменения в области наибольшей частотной чувствительности слуха. Примером является пороговая зависимость изменения группового времени запаздывания (ГВЗ, $\tau_{гр}= d\varphi(\Omega)/\Omega)$) от частоты F (рис. 1.9,*б*).

Можно сказать, что при амплитудной модуляции синусоидального сигнала тоном фазы боковых составляющих АМ колебания оказывают влияние на ощущение звука, пока все три частоты (несущее колебание и боковые составляющие) располагаются в одной критической полосе слуха. Если же они расположены в смежных частотных группах, то фазовые соотношения спектральных составляющих на слуховое ощущение никакого влияния не оказывают. Понятие критической полосы слуха будет пояснено позже.

Все же следует отметить, что временные свойства слуха являются наименее изученными.

1.3. Критические полосы слуха

Если в качестве полезного сигнала выступает тон, а в качестве маскирующего – узкополосный шум, центральная частота которого равна частоте тонального сигнала, и если полоса частот маскирующего шума ΔF расширяется, то при достижении некоторого значения $\Delta F = \Delta F_{кр}$ величина порога слышимости тона перестанет изменяться. Это значение полосы $\Delta F_{кр}$ и называется *критической полосой (частотной группой) слуха*. Оба понятия тождественны. Значение $\Delta F_{кр}$ зависит от средней частоты $F_{ср}$ (рис. 1.23, точки – экспериментальные данные). На частотах ниже 500 Гц ширина час-

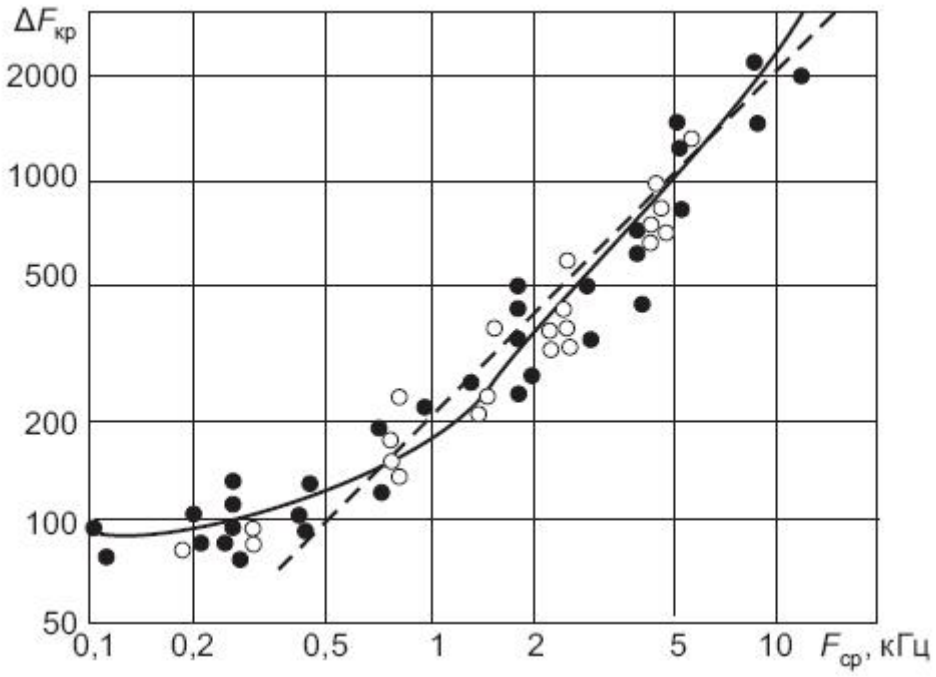


Рис. 1.23. Зависимость ширины критической полосы слуха от средней частоты. Точки на рисунке – экспериментальные данные разных авторов

тотной группы постоянна и составляет около 100 Гц. В области частот выше 500 Гц она возрастает пропорционально частоте, при этом $\Delta F_{кр} = 0{,}2F_{ср}$. В диапазоне частот от 20 Гц до 15,5 кГц размещаются 24 частотные группы слуха с шириной от 100 Гц в низкочастотной части и до 4...5 кГц в высокочастотной части спектра.

Слух сравнивает полезный сигнал и мешающий шум по интенсивности в пределах критических полос слуха, оценивая порог слышимости Полезный сигнал, например тон частотой 1000 Гц, на фоне мешающего шума воспринимается лишь тогда, когда его уровень ниже уровня мешающего шума в критической полосе слуха на 4 дБ.

В международных документах границы и средние частоты критических полос слуха соответствии данным табл.1.2. Критические полосы слуха не зависят от уровня интенсивности шума. При воздействии широкополосного шума слух как бы превращает сплошной спектр в дискретный. Такой спектр состоит из конечного числа составляющих по числу критических полос слухового анализатора.

Если ширина спектра узкополосного шума меньше ширины критической полосы слуха, то уровень громкости в этой полосе определяется лишь общей энергией шума и совершенно не зависит от характера распределения интенсивности в полосе. Она может быть распределена равномерно или сосредоточена в части полосы или быть в виде одного тона.

Таблица 1.2

Критические полосы слуха

Номер критической полосы слуха (частотной группы слуха)	Средняя частота критической полосы слуха, $F_{СР}$, Гц	Верхняя граничная частота критической полосы слуха, Fв, Гц	Ширина критической полосы слуха, $\Delta F_{КР}$, Гц
1	50	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150
9	1000	1080	160
10	1170	1270	190
11	1370	1480	210
12	1600	1720	240
13	1850	2000	280
14	2150	2320	320
15	2500	2700	380
16	2900	3150	450
17	3400	3700	550
18	4000	4400	700
19	4800	5300	900
20	5800	6400	1100
21	7000	7700	1300
22	8500	9500	1800
23	10500	12000	2500
24	13500	15500	3500
25	19500	-	-

Зависимость ширины критической полосы слуха от средней частоты $F_{ср}$ хорошо аппроксимируется формулой [1.7;1.8;1.9;1.10]

$$\Delta F_{КР}(F_{СР}) \approx 25+\exp[\ln(75)+0{,}69\cdot\ln(1+1{,}4\cdot F_{СР}^{2})]$$

$$\text{или } \Delta F_{КР}(F_{СР}) \approx 25+75\cdot(1+1{,}4\cdot F_{СР}^{2})^{0{,}69},$$

где $F_{СР}$ в кГц, $\Delta F_{КР}$ в Гц. Обе аппроксимации дают одинаковые результаты.

Каждой критической полосе слуха соответствует расстояние на базилярной мембране около 1,3 мм.

1.4. Критические полосы слуха и высота тона

Напомним, что границы критических полос слуха также были получены на основе ощущений, то есть они также образуют шкалу ощущений, для которых частота является параметром раздражения. Кривая, представленная на рис. 1.24, свидетельствует о том, что между высотой тона и шкалой,

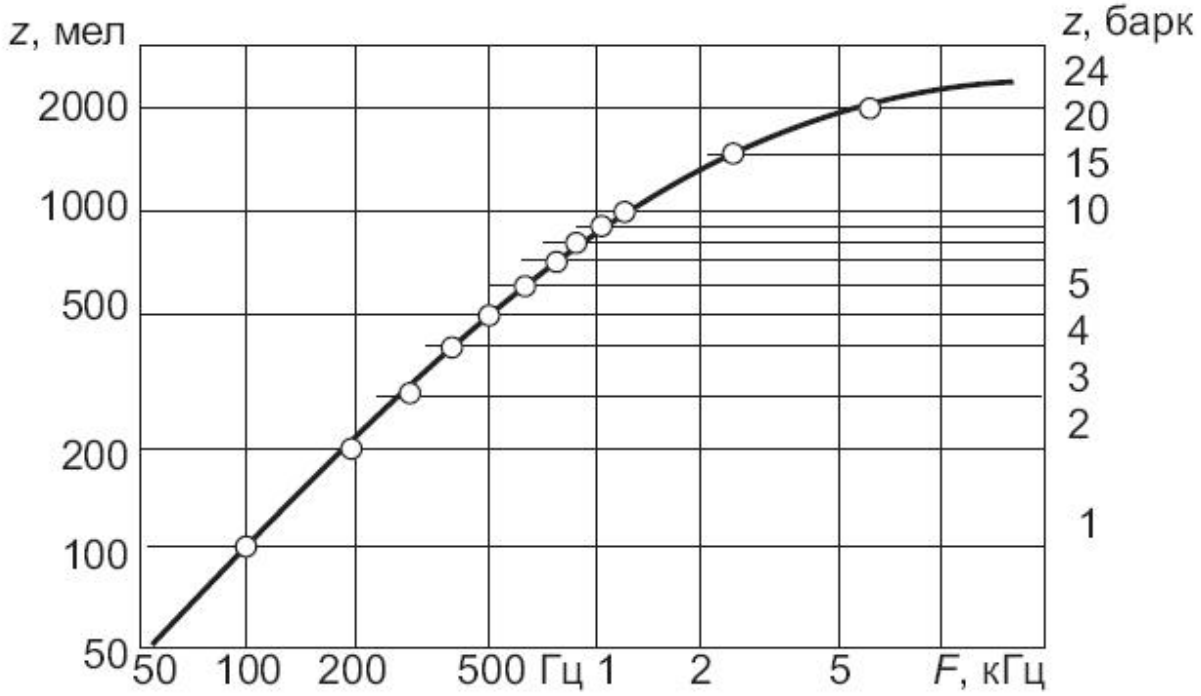


Рис. 1.24. Высота тона *z* и шкала частотных групп в зависимости от частоты

образованной тесно примыкающими друг к другу критическими полосами слуха, существует очень тесная связь. На рис. 1.24 по оси ординат слева отложена высота тона в *мелах*, по оси абсцисс – частота в логарифмическом масштабе. На правом краю данного графика также в логарифмическом масштабе, но со сдвигом в две декады нанесена шкала, образованная 24 сомкнутыми частотными группами. Если теперь по оси ординат откладывать эти 24 значения, а по оси абсцисс – верхние частоты соответствующих частотных групп (табл. 1.2), то получим точки, располагающиеся на кривой. Это значит, что высоту тона можно определить также и путем измерения порогов слышимости при маскировке по частотным группам и, что увеличение частоты на одну частную группу $\Delta F_{КР}$ всегда приводит к возрастанию соответствующей высоты тона *z* на 1 барк. По аналогии с частотной группой такой прирост высоты тона назовем *тональной группой*. Тональная группа играет весьма важную роль в ощущении силы звука.

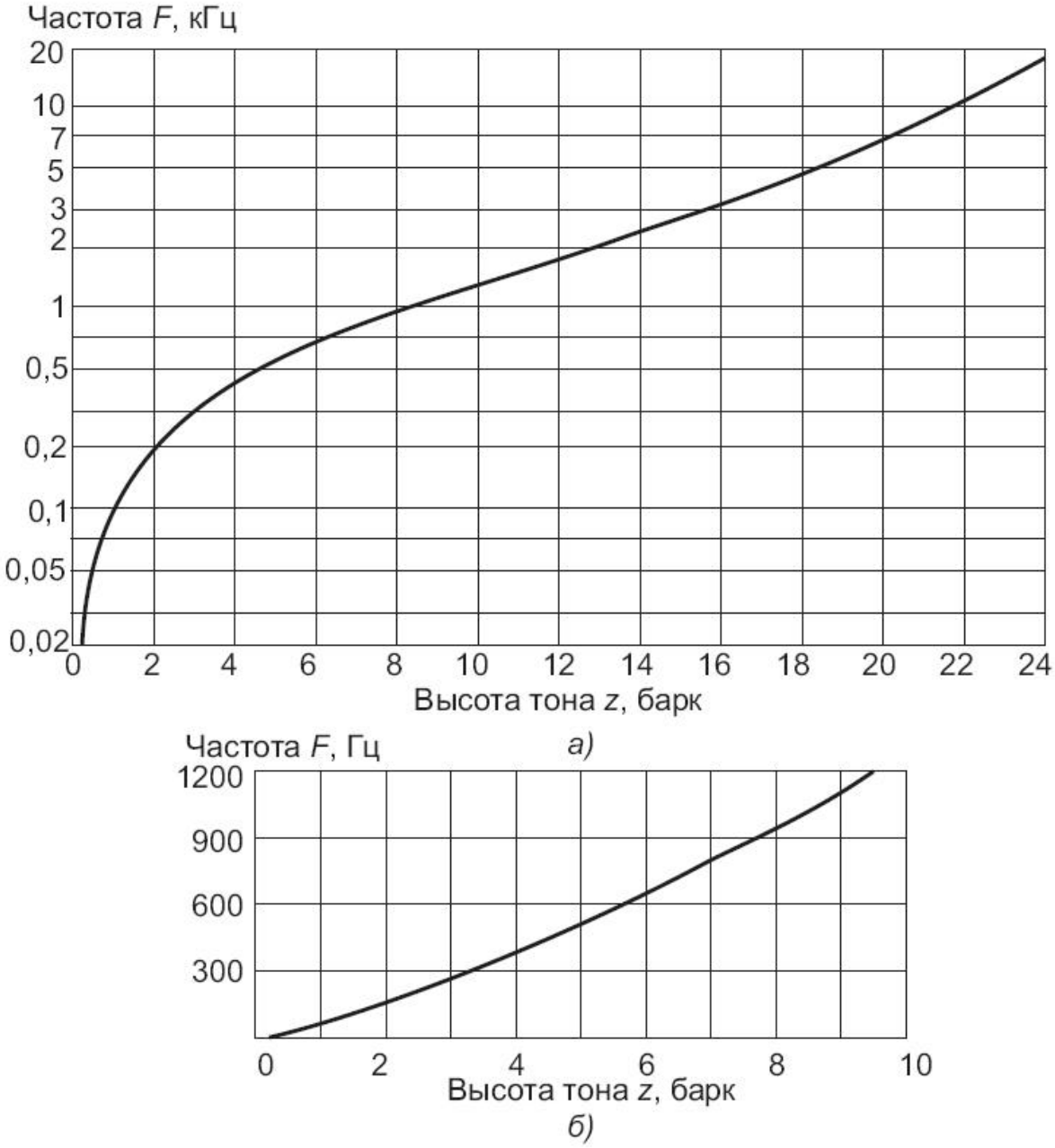


Рис. 1.25. Зависимость частоты от высоты тона: *а* – частота отложена в логарифмическом масштабе; *б* – частота отложена в линейном масштабе

Одним из первых, кто занимался измерениями силы звука, был Баркгаузен. В его честь была названа единица измерения высоты тона – *барк.* Увеличению частоты на одну частотную группу соответствует возрастание высоты тона на один барк. Напомним, что 1 барк равен 100 *мелам*. Зависимость между высотой тона z в барках и частотой еще раз показана на рис. 1.25,*а* и *б*. По оси абсцисс отложены в линейном масштабе значения высоты тона в барках, по оси ординат частота. Но на рис. 1.25,а она отложена в логарифмическом масштабе, а на рис. 1.25,*б* – в линейном. Из первой части этого рисунка следует, что на частотах выше 500 Гц (выше 5

барк) частота и высота тона связаны логарифмической зависимостью, а на частотах ниже 500 Гц (рис. 1.25,*б*) – уже линейной зависимостью.

Данная кривая (рис. 1.25) хорошо аппроксимируется зависимостью вида

$$z = 13{,}3 \cdot arctg(0{,}76 \cdot F) + 3{,}5 \cdot arctg(F/7{,}5)^2,$$

где частота F в кГц, высота тона z в барках. Возможны и другие аппроксимирующие функции, их аналитическая запись и ход показаны на рис. 1.26.

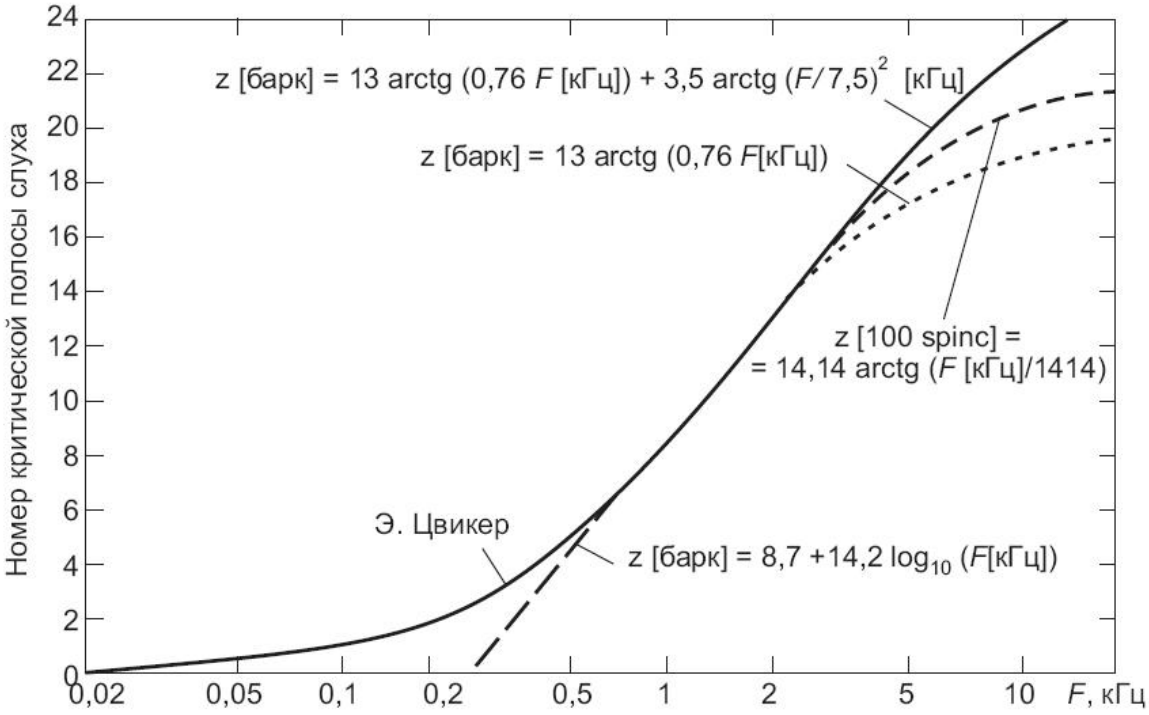


Рис. 1.26.Аппороксимирующие функции, устанавливающие связь между частотой F, кГц, и высотой тона z, барк (*а*) и шкалы слуховых ощущений (*б*)

Подробнее о шкалах, применяемых для оценки слуховых ощущений, связанных с изменением частоты и уровня тонального сигнала, см. в разд.1.8.

Номер критической полосы слуха в Барк-шкале может быть найден как

$$z_{кр} \approx [28{,}81/(1 + 1960/F] - 0{,}53,$$

где частота F в Гц, если получаемый по данной формуле результат $z_{1кр}<2$, то уточненное значение будет равно $z_{кр} = 0{,}15(2 - z_{1кр})$, если же получаемый результат оказывается большим 20,1, тогда уточненное значение может быть найдено как $z_{кр} = 0{,}22(z_{1кр} - 20{,}1)$.

И еще одно важное соотношение, позволяющее рассчитать для любого значения высоты тона z, заданной в барках, ширину критической полосы слуха $\Delta F_{КР}$ в Герцах

$$\Delta F_{кр} = 52548/(z^2 - 52,56z + 630,39).$$

1.5. Шкалирование слуховых ощущений

Итак, при воздействии тона слуховое ощущение изменения высоты связано с изменением частоты тона, а изменение громкости – с изменением уровня сигнала.

Для шкалирования ощущений, связанных с изменением частоты синусоидального сигнала, исследователями предложено несколько шкал: *SPINC*-шкала, Мел-шкала, Барк-шкала, *ERB*-шкала. Вернемся к рис. 1.26, чтобы пояснить это явление более подробно.

При построении *SPINC*-шкалы экспертам (слушателям) предъявлялись два синусоидальных сигнала (чистых тона) разной частоты с коротким временным интервалом между ними. Эксперт должен был определить минимальное расстояние между ними по частоте, при котором они разделяются (по ощущению высоты) или нет. Этот эксперимент позволяет определить разрешающую способность слуха по частоте в области слышимых звуков. Пороговое значение при данном опыте составляет около 1Гц для частот ниже 500 Гц, на более высоких частотах – около 2%. Например, если частота тона составляет 10000 Гц, то разрешение по частоте равно здесь 10,19 Гц. Величина порога зависит от уровня громкости звука, например для тона превышающего абсолютный порог слышимости от 5 до 10 дБ разрешающая способность слуха по частоте ухудшается.

Иной метод для оценки разрешающей способности слуха по частоте использовал Э.Цвикер. Для оценки порогов различимости по частоте он использовал, как мы уже знаем, частотно-модулированные сигналы: испытательный синусоидальный сигнал определенной частоты (его величина менялась в пределах области слышимых звуков) модулировался по частоте тоном, при этом оценивалось минимальное значение девиации частоты, замечаемое слухом. Частота модуляции составляла 4 Гц, что соответствует, как это было выяснено, наибольшей разрешающей способности слуха. Э. Цвикер установил, что для синусоидальных сигналов на частотах ниже 500 Гц порог различения по частоте составляет 3,6 Гц, а на более высоких частотах – 0,7%.

Е.Терхард [1.15] в 1988 году, аппроксимируя экспериментальные данные, полученные первым способом, предложил использовать следующее выражение для расчета разрешающей способности слуха по частоте

$$\Delta F_{П} = 1 + 0,5F^2,$$

где F – частота в кГц, $\Delta F_{п}$ – разрешающая способность слуха по частоте, в Гц. Исходя из этого выражения, он предложил шкалу для оценки величи-

ны слухового ощущения, вызванного изменением частоты, а также и название единицы измерения в данной шкале – *spinc* (*spectral pitch increment*), в этой шкале пороговое значение равно 1 *spinc*. Вышеприведенная формула была использована им для построения *SPINC*-шкалы, аналитическая форма записи которой представлена ниже

$$\Phi(F) = 1414 \cdot \mathrm{arctg}(F/1414),\ spinc,$$

где частота F в Гц, а значение Ф в spinc. Вид этой функции представлен на рис. 1.26. В более поздних работах интервалы (шаги) этой шкалы были уточнены. Заметим, что частоте 1000 Гц соответствует 1000 мел или 8,5 барк или 870,4 *spinc*. Следовательно, интервалу в 870,4 *spinc* соответствует изменение частоты от 0 до 1000 Гц. Частоте 500 Гц – 480,6 spinc; частоте 2000 Гц – 1351 *spinc*; частоте 5000 Гц – 1831,5 *spinc*. Частоте 16000 Гц соответствует – 2096,5 *spinc*, интервалу в 2096,5 *spinc* соответствует частотный интервал 0…16000 Гц. Если известно изменение слухового ощущения в *SPINC*-шкале ΔФ, то соответствующее ему изменение частоты может быть найдено из выражения

$$\Delta F = 2828 \cdot tg\Delta\Phi \cdot \frac{1+(F_1/1414)^2}{1-(\Delta\Phi/2828)^2 \cdot (F_1/1414)^2},\ \text{Гц},$$

где F_1 – частота в Гц, ΔФ, в *spinc*.

С. Стивенс, исследуя эту проблему, ввел для шкалирования ощущения высоты тона уже известную нам ранее Мел-шкалу, интервалом которой является единица измерения мел. Он предложил следующее выражение, связывающее частоту и высоту тона

$$z(F) = 3322(\log(1000+F)-3),$$

где z – высота тона в мелах, F – частота в Гц. Позже интервалы этой шкалы были уточнены Э. Цвикером. С этой точки зрения Мел-шкалы, предложенные С. Стивенсом и Э. Цвикером, отличаются друг от друга. Напомним, что Мел-шкала Э. Цвикера базируется на уже известной нам Барк-шкале, последняя, как известно, однозначно связана с процессами, протекающими на базилярной мембране при изменении частоты воздействующего тона и с критическими полосами слуха (рис. 1.26 и 1.27).

Е.Терхард позже предложил формулу для расчета ширины критических полос слуха вида

$$\Delta F_{кр} = 86 + 0{,}055F^{1{,}4},\ \text{Гц},$$

где F – частота, также в Гц. В соответствии с этой формулой ширина критических полос слуха зависит от частоты, она составляет около 100 Гц на низких частотах и возрастает с повышением частоты: для средней частоты 10500 Гц формула дает значение равное 2500 Гц, и равна 9000 Гц на частоте 11500 Гц, при этом средняя частота располагается несимметрично. Более точной следует считать Барк-шкалу, которая имеет 24 ступени в полосе частот от 0 до 16000 Гц. Значения частоты F, кГц, и высоты тона в Барк-шкале связаны выражением

$$z = 13{,}3 \cdot arctg(0{,}76 \cdot F) + 3{,}5 \cdot arctg(F/7{,}5)^2 .$$

И, наконец, так называемая *ERB*-шкала (*ERB - equivalent rectangular bandwith*), также связывающая величину ощущения с изменением частоты, была предложена в 1990 году (*Glasberg* и *Moore*). Ее получение основано на следующей идее. Интервалы (шаги) в этой шкале соответствуют ширине полосы частот фильтра с идеальной прямоугольной формой амплитудно-частотной характеристики, который вырезает из белого шума точно такое же количество энергии, как и реальный фильтр базилярной мембраны. При этом средняя частота этого фильтра соответствует точке базилярной мембраны с наибольшим возбуждением. Очевидно, что это более узкие по полосе частот фильтры, ее ширина 1 *ERB*, для ее оценки ширины полосы частот такого идеального фильтра предложена следующая формула

$$\Delta F_{ERB} = 24{,}7(4{,}37F + 1),$$

где ΔF_{ERB} в Гц, а частота F в кГц. С помощью данного выражения получают ERB-шкалу, в которой цена одного шага составляет 1 *ERB*. Обе шкалы (*ERB*-шкала и Барк-шкала) дают почти совпадающие результаты на частотах превышающих 1000 Гц, ниже 1000 Гц ширина частотных групп оказывается шире полосы частот эквивалентного фильтра с прямоугольной формой АЧХ. Для частоты 1000 Гц оценки ощущения в *ERB*-шкале и Барк-шкале практически совпадают. В общей сложности *ERB*-шкала в полосе частот имеет 40 шагов, в то время как Барк-шкала лишь 24. Можно сказать, что в области низких частот шкалирование ощущений высоты тона с помощью *ERB*-шкалы является более точным.

Уравнение, устанавливающее связь между *ERB* оценкой B_{ERB} и частотой F, имеет вид

$B_{ERB} = 21{,}4 \cdot \lg(4{,}37F + 1)$, где B_{ERB} в *ERB* единицах, F – частота в кГц,

или $B_{ERB}=11{,}17268\lg(1+\frac{46{,}06538F}{F+14678{,}49})$, где B_{ERB} в ERB единицах, частота F в Гц.

Для обратного пересчета можно воспользоваться формулой вида

$$F=\frac{676170{,}4}{47{,}06538\cdot e^{0{,}08950401B_{ERB}}}-14678{,}49.$$

1.6. Одновременная маскировка

Маскировка – это есть изменение порога слышимости одного сигнала в присутствии другого мешающего звука. Мешающими звуками могут быть шум, созвучие, чистый тон. При оценке маскировки регистрируются уровни полезного сигнала при наличии и отсутствии мешающего сигнала. Очевидно, что кривые порога слышимости маскируемого сигнала всегда лежат выше абсолютного порога слышимости.

Если оба звука воздействуют одновременно, то мы имеем дело с *одновременной маскировкой*.

Если воздействие обоих сигналов разнесено во времени, то имеет место *временная маскировка.*

В противоположность этому при разнесении источников звука в пространстве начинает работать механизм *бинауральной демаскировки*. Он проявляется в том, что на фоне общего разговора множества людей можно «настроившись» на интересующего нас человека прослушать его разговор. Это сделать гораздо проще, когда говорящие люди разнесены в пространстве относительно слушающего человека. Этот феномен получил название «эффекта вечеринки» (*Cocktail Party Effect*).

Маскирующее действие различных звуков выявляют путем оценки порога слышимости маскируемого сигнала в присутствии маскирующего (*masker*) сигнала. Для каждого уровня мешающего шума, созвучия, тона существует свой порог слышимости маскируемого сигнала. В отличие от абсолютного порога слышимости данный порог называют относительным.

Пороги слышимости при одновременной маскировке

Маскировка тона маскируемого тоном. Пусть частота маскирующего тона составляет $F_м$, а его уровень $N_м$. Изменяя частоту измерительного (маскируемого тона) в пределах области слышимых частот и определяя для каждого значения его частоты относительный порог слышимости, можно получить кривую порога слышимости тона при наличии маскировки. Форма такой кривой дана на рис. 1.28 (пунктирная линия). По оси орди-

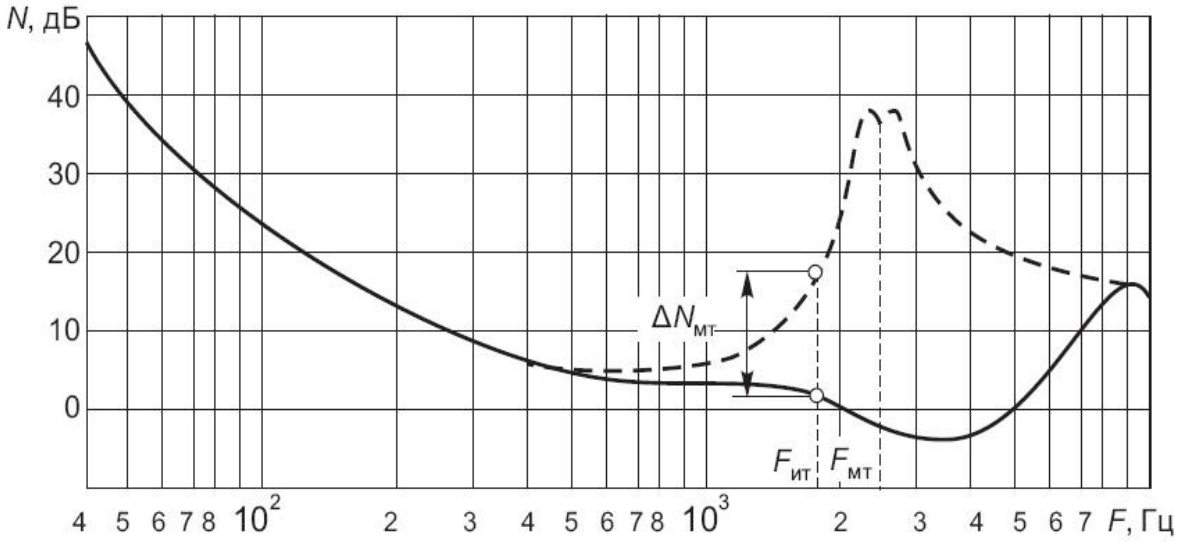


Рис. 1.28. Кривая полога слышимости тона при маскировке чистым тоном

нат отложены уровни интенсивности звука N, дБ, по оси абсцисс – значения частоты F, в Гц. Нижняя кривая представляет собой порог слышимости в тишине. Частота маскирующего тона равна $F_{МТ}$ = 2400 Гц, а его уровень интенсивности 60 дБ. Если значение частоты маскируемого (измерительного) тона равно $F_{ИТ}$ = 1400 Гц, то из данной кривой можно найти величину $\Delta N_{МТ}$, показывающую повышение полога слышимости маскируемого сигнала по сравнению со значением его абсолютного порога слышимости. Величина $\Delta N_{МТ}$ может служить количественной мерой маскирующего действия тона частотой 1400 Гц в присутствии тона частотой 2400 Гц.

Для удобства оценки величины маскировки часто значение абсолютного порога слышимости принимают нулевой уровень и рассматривают лишь это превышение $\Delta N_{МТ}$. В качестве примера на рис. 1.29. представлены семейства кривых маскировки для разных частот и уровней маскирующих тонов. По оси абсцисс отложена частота маскируемого тона, по оси абсцисс – величина маскировки $\Delta N_{МТ}$, дБ. Параметром каждой кривой семейства является уровень маскирующего тона $N_{МТ}$, дБ.

Все представленные здесь семейства кривых маскировки имеют следующие общие особенности:

-наибольшая маскировка наблюдается, когда частоты обоих тонов близки;

- маскировка увеличивается с ростом уровня маскирующего сигнала, несимметричность кривых маскировки также растет с увеличением уровня маскирующего сигнала;

- маскировка в сторону низких частот проявляется значительно слабее, чем в сторону высоких, более сильное маскирующее действие в сторону высоких частот, по-видимому, обусловлено возникновением в слуховой системе субъективных гармоник маскирующего тона, вызванных возбуж-

дением соответствующих участков базилярной мембраны и повышающих вследствие этого порог слышимости маскируемого сигнала вблизи этих об-

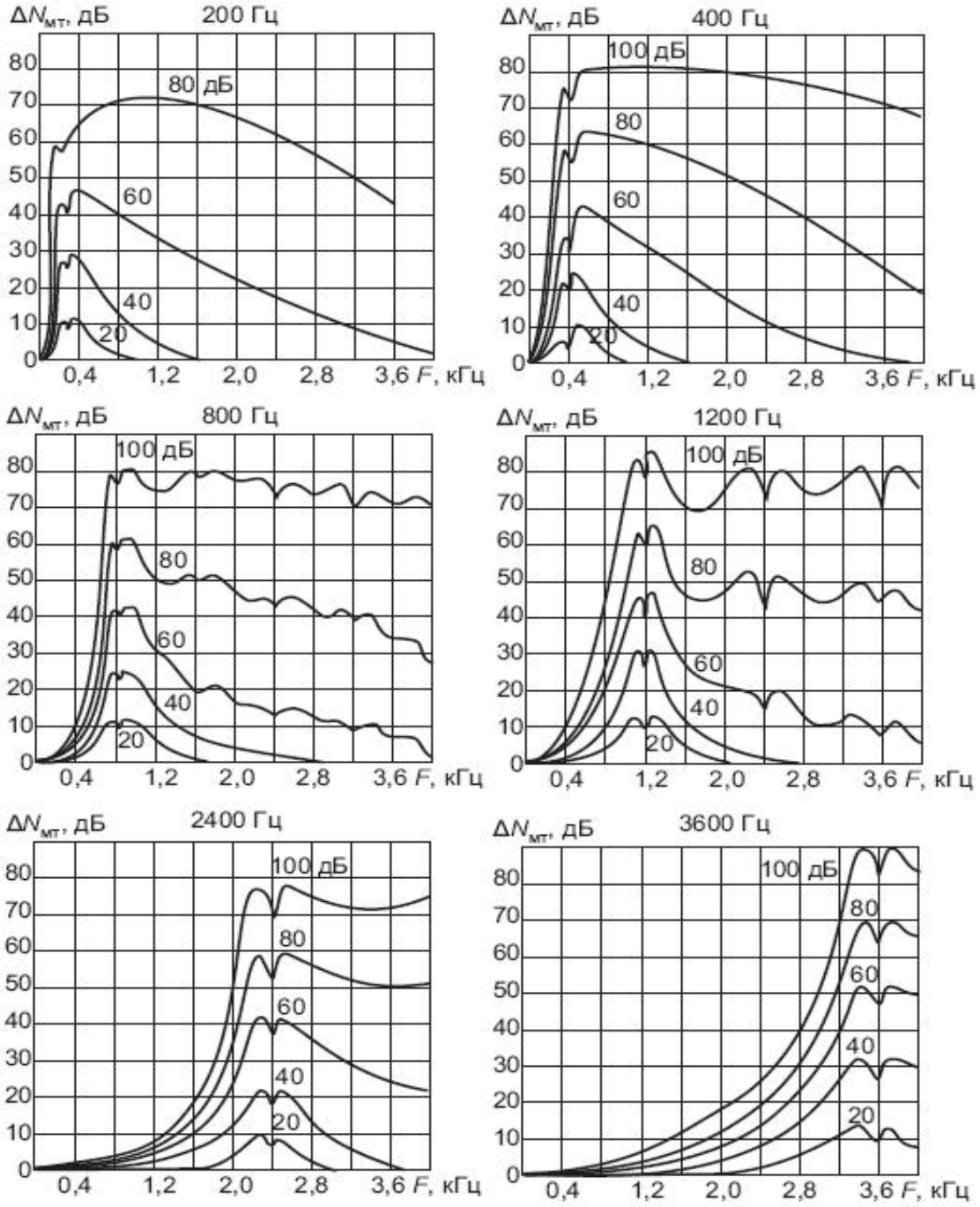


Рис. 1.29. Кривые маскировки тонов маскируемых тональными сигналами разных значений частот и уровней: *а* – частота маскирующего тона 200 Гц; *б* – частота маскирующего тона 400 Гц; *в*, *г*, *д* – то же самое, но соответственно для маскирующих тонов с частотами 800, 1200, 2400 и 3600 Гц

ластей. Это подтверждается и наличием впадин на кривых маскировки, являющихся следствием возникновения биений между субъективными гармониками маскирующего тона и маскируемым сигналом. Кривые маски-

ровки тона тоном частотой 1000 Гц, взятые из другого источника, представлены на рис. 1.30. Здесь $N_М$ – уровень маскирующего тона, дБ.

Маскировка тона узкополосным шумом. Если в качестве маскирующего сигнала используется узкополосный шум, то кривые порога слышимости тона имеют вид, показанный на рис. 1.31,*а*. В качестве маскирующе-

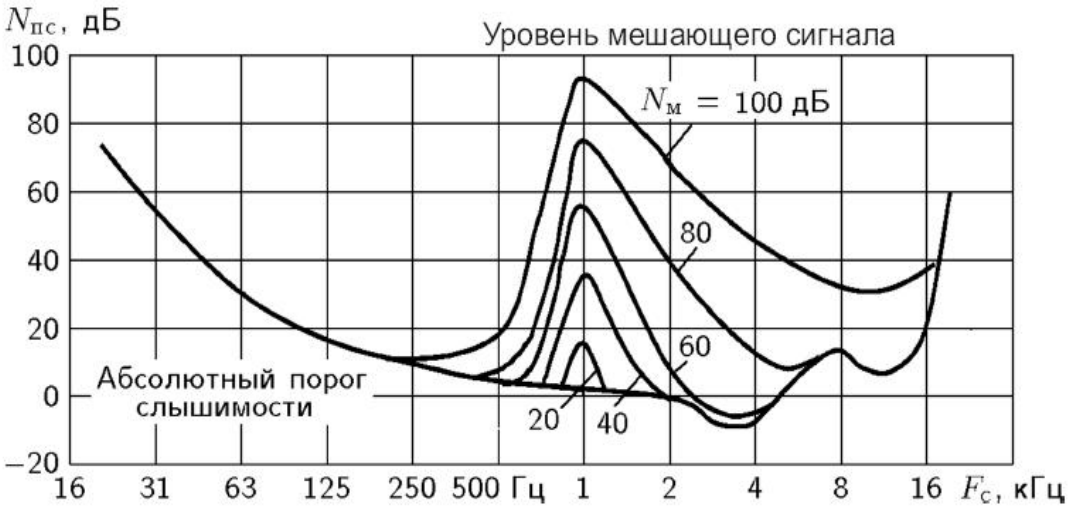


Рис. 1.30. Кривые порога слышимости тона маскируемого тоном частотой 1000 Гц ; нижняя кривая – абсолютный порог слышимости тона

го сигнала здесь взят узкополосный шум со средней частотой 1000 Гц, полосой частот 160 Гц и уровнями интенсивности $N_{МУШ}$ = 20, 40, 60, 80 и 100 дБ. Параметром каждой кривой является уровень $N_{МУШ}$, дБ. По оси ординат отложены уровни N, дБ, тона на пороге слышимости, по оси абсцисс – частота тона, в Гц. Все кривые имеют явно выраженный максимум на частоте 1000 Гц, его значение лежит на 4 дБ ниже соответствующего уровня маскирующего сигнала. Кривые маскировки несимметричны: быстро спадают в сторону низких частот (здесь они почти параллельны друг другу) и значительно более медленно в сторону высоких частот. При повышении уровня маскирующего сигнала кривые маскировки расширяются и спадают в сторону верхних частот тем медленнее, чем выше уровень маскирующего шума. Следовательно, область частот, где проявляется маскировка, расширяется, тем больше, чем выше уровень маскирующего сигнала. Можно сказать, что высокие частоты слабо маскируют более низкие частоты. В свою очередь низкие частоты оказывают существенно большее маскирующее действие, охватывающее существенно более широкую область частот.

На рис. 1.31,*б* представлены кривые маскировки тона для узкополосных

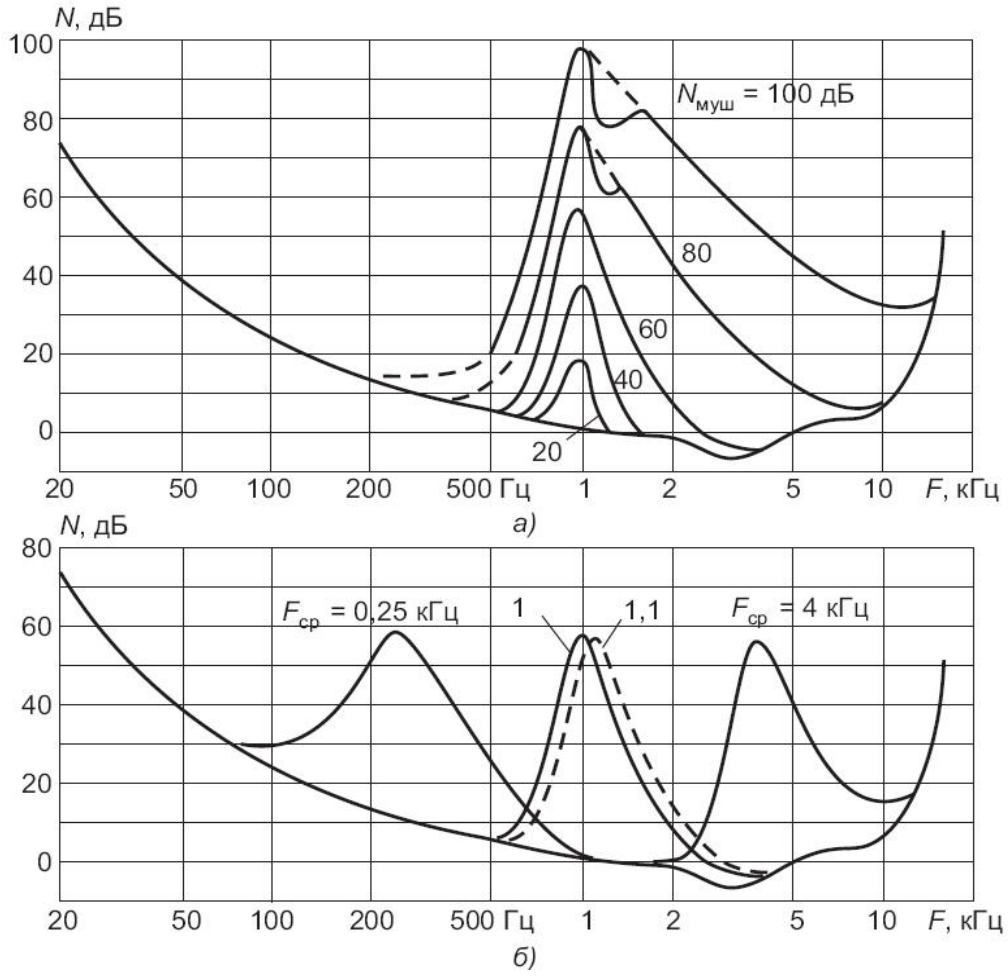


Рис. 1.31. Кривые маскировки тона узкополосным шумом
с различными средними частотами

шумов со средними частотами 0,25, 1,0, 1,1 и 4 кГц. Во всех случаях значение $N_{МУШ}$ = 50 дБ. В точках максимумов кривые маскировки достигают уровней, лежащих ниже уровня маскирующего узкополосного шума на 4 дБ. Незначительное изменение средней частоты шума приводит к небольшому смещению кривых маскировки параллельно самим себе. На низких частотах область частот, где проявляется маскировка, сравнительно уже (носит локальный характер), чем на высоких частотах, где она охватывает уже значительную область частот.

Маскировка тона белым шумом. Большой интерес для практики представляют кривые порога слышимости тона при его маскировке широкополосным белым шумом. Семейство данных кривых представлено на рис. 1.32. Параметром каждой такой кривой является уровень интенсивности маскирующего белого шума $N_{МБШ}$, дБ. До частоты 500 Гц (после ответвления от кривой абсолютного порога слышимости) они идут параллельно оси абсцисс (горизонтально), на более высоких частотах (F > 500 Гц)– это возрастающие прямые. Крутизна нарастания у всех у них одина-

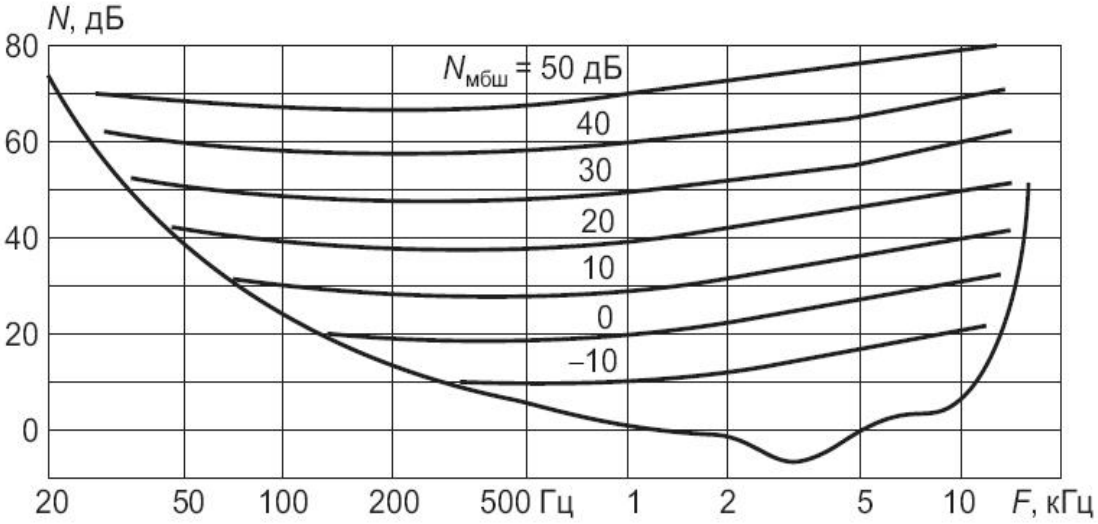


Рис. 1.32. Кривые порога слышимости тона при маскировке широкополосным белым шумом

кова: при возрастании частоты на октаву уровень порога слышимости повышается на 3 дБ. Как и ранее, степень маскировки зависит от уровня интенсивности маскирующего шума. При этом увеличение уровня маскирующего сигнала на 10 дБ вызывает повышение порога слышимости тона также на 10 дБ.

Форму этих кривых можно объяснить, если учесть, что слух реагирует не на общую мощность шума во всей полосе частот, а на его мощность в критических полосах слуха. На низких частотах ширина (F < 500 Гц) ширина критических полос слуха примерно одинакова. Одинакова, следовательно, и мощность шума в каждой из них, ибо спектральная мощность белого шума постоянна во всей полосе частот. На более высоких частотах (F > 500 Гц) ширина критических полос слуха растет с увеличением средней частоты $\Delta F_{КР} \approx 0{,}2F_{СР}$. Поэтому на частотах выше 500 Гц уровень интенсивности маскирующего шума в каждой из критических полос слуха возрастает с повышением их номера, что следует из выражения

$$N_{БШ} = 10\lg(S\cdot\Delta F/S_0\cdot 1\ Гц) = 10\lg(S/S_0) + 10\lg(\Delta F/1\ Гц)$$
$$или\ N_{БШ} = N + 10\lg(\Delta F/1\ Гц),$$

где $N_{БШ}$ – уровень интенсивности белого шума, ΔF – полоса частот белого шума; N – уровень спектральной плотности интенсивности белого шума (мощности, приходящейся на 1 Гц), в дБ; S_0 – спектральная плотность интенсивности шума на нулевом уровне 10^{-12} Вт/м^2.

Итак, мощность шума, приходящаяся на каждую критическую полосу слуха на частотах выше 500 Гц, растет с увеличением средней частоты шума, и, как следствие этого, повышается порог маскировки. Из выше приведенного выражения следует, что при каждом повышении частоты на октаву

порог слышимости повышается на 3 дБ в области, где частота $F_{СР}$ > 500 Гц. Это свойство слуха, как показал впервые Г.Флетчер, может быть использовано для уточнения ширины критических полос слуха. Если поддерживать уровень спектральной плотности шума постоянным и расширять полосу частот шума и каждый раз при новой полосе шума определять порог слышимости тона, который охвачен шумом с двух сторон, то порог слышимости тона будет нарастать до тех пор, пока полоса шума не станет равна критической полосе слуха. Дальнейшее расширение полосы шума не будет приводить к повышению порога слышимости.

Маскировка тона равномерно маскирующим белым шумом. В ряде акустических исследований необходимо иметь шум, который давал бы кривые маскировки, идущие параллельно оси абсцисс. Для этого нужно, чтобы уровень спектральная плотности шума оставался бы постоянным до частоты 500 Гц, а далее с ростом частоты, уровень его спектральной плотности интенсивности уменьшался бы со скоростью 3 дБ на октаву. Для этого белый шум требуется пропустить через фильтр с соответствующей частотной характеристикой (рис. 1.33). Соответствующие такому шуму

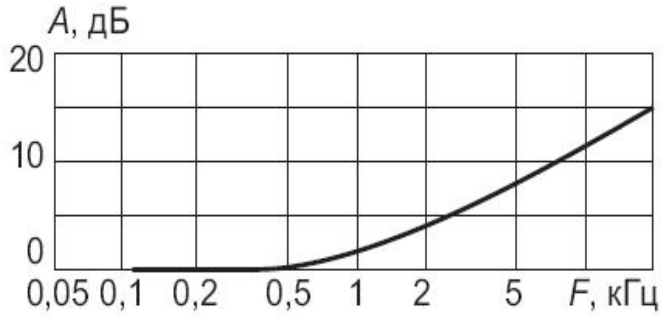


Рис. 1.33. Частотная характеристика затухания *A*, дБ, фильтра для получения равномерно маскирующего шума из белого шума

кривые маскировки тона показаны на рис. 1.34. Параметром каждой такой кривой является уровень интенсивности равномерно маскирующего белого шума $N_{РМШ}$ на частотах ниже 500 Гц или исходного белого шума. Все кривые проходят примерно на 18 дБ выше уровня спектральной плотности исходного белого шума. В данном случае уровень порога слышимости тона $N_{ПС}$ находится из выражения

$$N_{ПС} = N + 18 \text{ дБ},$$

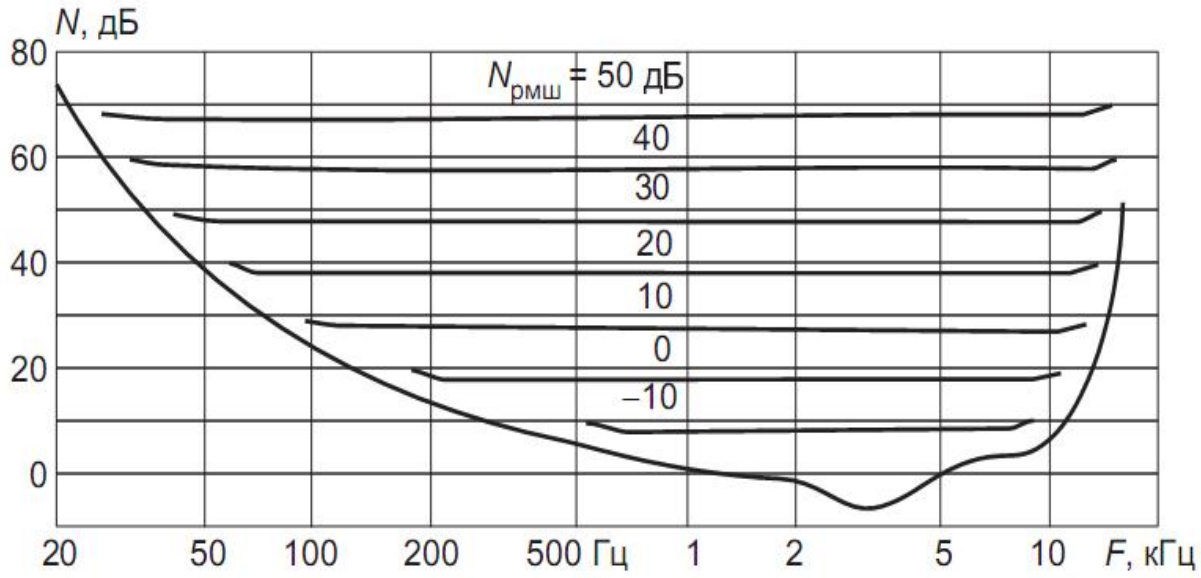


Рис. 1.34.Кривые порога слышимости тона при маскировке равномерно маскирующим шумом

где N – уровень спектральной плотности белого шума, из которого путем фильтрации (рис. 1.33) получен равномерно маскирующий шум.

Маскировка тона реальными звуковыми сигналами. Реальные звуковые сигналы имеют гораздо более сложные структуре и динамике спектры и, как следствие этого, и более сложные кривые маскировки. На рис. 1.35 показаны кривые маскировки тона звуками скрипки различного уровня, издающей самый низкий тон частотой 195 Гц (*а*) и высокий тон (*б*). Звук скрипки состоит из основного тона и множества слабо убывающих по уровню обертонов. В силу этой причины кривая порога слышимости тона при маскировке звуком скрипки почти горизонтальна в широком диапазоне частот, начиная с частоты основного тона. Эти кривые пересекаются узкими участками биений, когда частота измерительного тона очень близко совпадают с частотами обертонов. Зависимости, показанные на рис. 1.35 близки по форме. Во всем диапазоне частот выше частоты основного тона кривые маскировки располагаются значительно выше кривой абсолютного порога слышимости (нижняя кривая). Высокие частоты слабо маскируют низкие, ибо кривые маскировки резко обрываются в сторону низких частот и наоборот: низкие частоты оказывают сильное маскирующее действие на высокие частоты. Именно по этой причине в смешанных хорах мужских голосов всегда меньше, чем женских, а в симфоническом оркестре виолончелей намного меньше, чем скрипок. Не последнюю роль здесь играет и тот факт, что акустическая мощность источников звука высоких частот, как правило, существенно меньше, чем мощность источников низких частот.

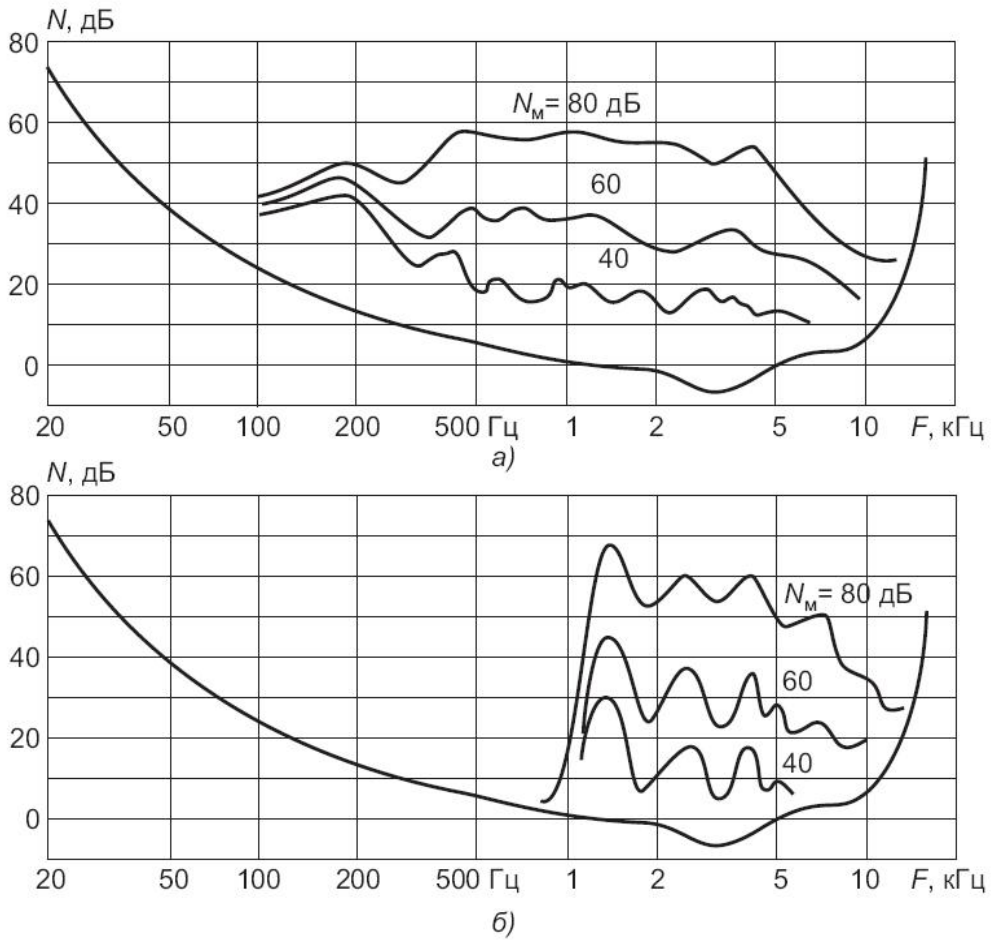


Рис. 1.35. Кривые маскировки тона низким (*а*) и высоким (*б*) звуками скрипки различного уровня

Моделирование одновременной маскировки

Одновременная маскировка проявляется по-разному в зависимости от особенностей спектров сигналов. При разработке алгоритмов компрессии цифровых аудиоданных учитывается различие маскировки внутри (*intra-band masking*) и вне (*extra-band masking*) критической полосы (частотной группы) слуха.

Маскировка внутри критической полосы слуха (***intra-band masking***). Она оценивается с помощью коэффициента маскировки $\Delta N_{\text{M.}}$ Этот коэффициент имеет разное значение в зависимости от того маскирует ли тон шум (ΔN_{M1}), или, наоборот (ΔN_{M2}). В первом случае берется шум, охватывающий равномерно тон с шириной полосы равной частотной группе слуха и определяется порог слышимости шума маскируемого тоном (*tone masking noise),* во втором - шум с полосой частот, равной частотной группе слуха, маскирует тон (*noise masking tone*).

Маскировка тона шумом *(ΔN_{M2}).* Под коэффициентом маскировки ΔN_{M2} понимается разность уровней чистого тона на пороге слышимости и

маскирующего шума равномерно охватывающего этот тон и имеющего полосу частот в одну частотную группу. Величина коэффициента маскировки ΔN_{M2} до частоты 500 Гц минимальна и равна около минус 2 дБ. С возрастанием частоты в области частот, где ширина критических полос слуха растет, величина коэффициента маскировки ΔN_{M2} уменьшается, приближаясь на самых верхних частотах к значению минус 6 дБ.

Коэффициент маскировки тона шумом ΔN_{M2} аппроксимируется функцией

$$\Delta N_{M2} \approx -2{,}0 - 2{,}05 \cdot \mathrm{arctg}(F/4) - 0{,}75 \cdot \mathrm{arctg}(F^2/2{,}56)\ ,$$

где: F – частота в кГц, а ΔN_{M2} – в дБ. Зависимость ΔN_{M2} от частоты представлена на рис. 1.36,*а*. По оси абсцисс отложены значения частоты тона F, кГц;

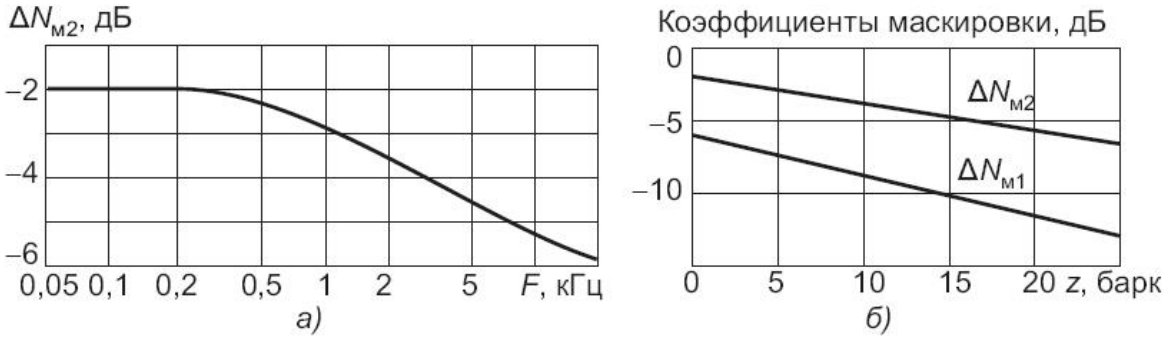


Рис. 1.36. Зависимость коэффициента маскировки от частоты (а) и высоты тона (б, прямая 1 – значения ΔN_{M2}, прямая 2 – значения ΔN_{M1})

по оси ординат - значения ΔN_{M2} в дБ. Во всех случаях тон становится слышимым даже, если его уровень меньше уровня маскирующего шума. Можно воспользоваться для расчета величины ΔN_{M2} и более простым выражением вида:

$$\Delta N_{M2} \approx -1{,}525 - 0{,}175 \cdot z - 0{,}5,$$

где ΔN_{M2-} в дБ, а z – в барках (рис. 1.36,*б*, прямая 1). Это выражение используется в стандарте *MPEG*-1 *ISO/IEC* 11172-3.

Маскировка шума тоном (ΔN_{M1}). Совершенно другая картина имеет место, когда маскирующим сигналом является чистый тон, а маскируемым - шум. Здесь явление маскировки проявляется значительно слабее. В качестве аппроксимирующей функции часто используется выражение вида [1.10]:

$$\Delta N_{M1} \approx -(15{,}5 + z)\ ,\ \text{дБ}.$$

Здесь z – высота тона, в барках. Зависимости ΔN_{M1} (нижняя кривая) и ΔN_{M2} (верхняя кривая) представлены на рис 1.36,*б*. Здесь по оси абсцисс отложены значения высоты тона в барках, по оси ординат - значения коэффициентов маскировки ΔN_{M1} и ΔN_{M2}, в дБ. Видно, что с увеличением z значения коэффициентов маскировки уменьшаются. При одном и том же значении z всегда ΔN_{M2} существенно превышает ΔN_{M1}. Известны и другие аппроксимации для кривых маскировки, при этом выражения, предлагаемые разными авторами для описания этого явления, имеют существенные расхождения.

Маскировка вне критической полосы слуха. (***extra-band masking***). Для аппроксимации кривых маскировки для этого случая предложено множество функций. Одной из наиболее распространенных является пара функций вида (по *Terhardt E.*, [1.9]):

$$B(z) \approx 10^{0,1\cdot[A1\cdot(Z-Z_{M})]} \qquad \text{при } z< z_M$$
$$B(z) \approx 10^{-0,1[A2\cdot(Z-Z_{M})]} \qquad \text{при } z>z_M$$

где $A1$=27 дБ/барк; $A2$=[24+230/F -0,2·N_M], дБ/барк; N_M-уровень маскирующего сигнала, дБ, z – высота тона маскируемого сигнала, в барках; z_M – высота тона маскирующего сигнала, в барках, $B(z)$ – относительное значение интенсивности маскируемого сигнала с высотой тона z, соответствующее порогу его слышимости в присутствии маскирующего сигнала с высотой тона z_M; $B(z)$ – нормированная кривая маскировки. Очевидно, что величина относительного порога слышимости имеет максимальное значение при $z = z_M$. Заметим, что часто кривые маскировки, учитывающие маскировку вне критической полосы слуха, называют «развертывающими функциями» («*Spreading Funktion*»).

Известны и другие аппроксимирующие функции табл. 1.3.

Таблица 1.3

Аппроксимирующие функции для расчета кривых маскировки, предложенные разными авторами

Номер литературного источника	Аппроксимирующее выражение
[1.9]	$B(z) \approx 10^{0,1\cdot[A1\cdot(z-z_{м})]}$ при $z< z_M$ $B(z) \approx 10^{-0,1[A2\cdot(z-z_{м})]}$ при $z< z_M$
[1.21]	$B(\Delta z)=\begin{cases} 17(\Delta z+1)-[0,4N_M(z_M)+6] & \text{при} \quad \Delta z\in(-3...-1), \\ [0,4N_M(z_M)+6]\Delta z & \text{при} \quad \Delta z\in(-1...0), \\ -17\Delta z, & \text{для} \quad \Delta z\in(0...1), \\ -(\Delta z-1)\cdot[17-0,15N_M(z_M)]-17 & \text{при} \quad \Delta z\in(1...8) \end{cases}$ где $\Delta z = z - z_M$, причем z – высота тона маскируемой компоненты, барк, z_M – высота тона маскера, барк; $N_M(z_M)$ – уровень интенсивности маскера, в дБ

Окончание табл. 1.3

Номер литературного источника	Аппроксимирующее выражение
[1.22]	$B(\Delta z) = A_1 + \left(\frac{D_1 + D_2}{2}\right)(\Delta z + A_2) - \left(\frac{D_1 - D_2}{2}\right)\sqrt{\gamma + (\Delta z + A_2)}$, причем $A_1 = \sqrt{\gamma \cdot D_1 \cdot \lvert D_2 \rvert}$, $A_2 = \frac{D_1 + D_2}{2} \times \sqrt{\frac{\gamma}{D_1 \cdot \lvert D_2 \rvert}}$, $D_1 = 27$ дБ/барк, $D_2 = -(22 - 0,2 \cdot N_M(z_M))$ дБ/барк, где D_1 и D_2 – соответственно крутизна переднего и заднего фланцев кривой маскировки; γ – коэффициент кривизны перехода от переднего к заднему фланцу «развертывающей функции»
[1.23]	$B(\Delta z) = 15,81 + 7,5 \cdot (\Delta z + 0,474) - 17,5 \times \sqrt{\left(1 + (\Delta z + 0,474)^2\right)}$ получено из [22] при условии, что $D_1 = 25$ дБ/Барк, $D_2 = -10$ дБ/барк, $\gamma = 1$
Стандарт *MPEG-1 ISO/IEC* 11172-3, психоакустическая модель 2	$B(\Delta z) = A_0 + 15,81 + 7,5(\Delta z + 0,474) - 17,5\sqrt{\left(1 + (\Delta z + 0,474)^2\right)}$, $A_0 = 8\min\left((\Delta z - 0,5)^2 - 2(\Delta z - 0,5), 0\right)$, для полосы психоакустического анализа
[1.24]	$B(\Delta z) = \begin{cases} N_M(z_M) + D_1 \cdot (\Delta z + 0,5), & \text{при} \quad \Delta z < -0,5 \\ N_M(z_M), & \text{при} \quad -0,5 \le \Delta z \le 0,5 \\ N_M(z_M) + D_2 \cdot (\Delta z - 0,5), & \text{при} \quad \Delta z > 0,5 \end{cases}$ где $D_1 = 27$ дБ/барк, $D_2 = -10$ дБ/барк
[1.25]	$B(\Delta z) = \begin{cases} D_1 \cdot \Delta z, & \text{для} \quad \Delta z < 0 \\ D_2 \cdot \Delta z, & \text{для} \quad \Delta z \ge 0 \end{cases}$, где $D_1 = 27$ дБ/барк, $D_2 = -\left(24 + \frac{230}{F} - 0,2 \cdot N_M(z_M)\right)$, дБ/барк.
[1.26]	$B(\Delta z) = \begin{cases} D_1 \cdot \Delta z, & \text{при} \quad \Delta z < 0 \\ D_2 \cdot \Delta z \cdot \left(\frac{1}{1 + \frac{\Delta z}{5}}\right), & \text{при} \quad \Delta z \ge 0 \end{cases}$, где $D_1 = 27$ дБ/барк, $D_2 = -24$ дБ/барк

В качестве примера на рис. 1.37 представлено семейство индивидуальных кривых маскировки (развертывающих функций) для различных уровней маскирующего тона: *а* – психоакустическая модель 1 стандартов MPEG *ISO/IEC* 11172-3 и 13818-3; *б* – психоакустическая модель 2 стандартов *MPEG ISO/IEC* 11172-3, 13818-3, 14496-3. Для удобства сравнения на рис.

1.38, 1.39 изображены развертывающие функции, полученные разными авторами, для трех значений уровня маскирующей компоненты $N_М$ =20 дБ, 60 дБ и 96 дБ. Видно, что все предлагаемые аппроксимации развертывающих

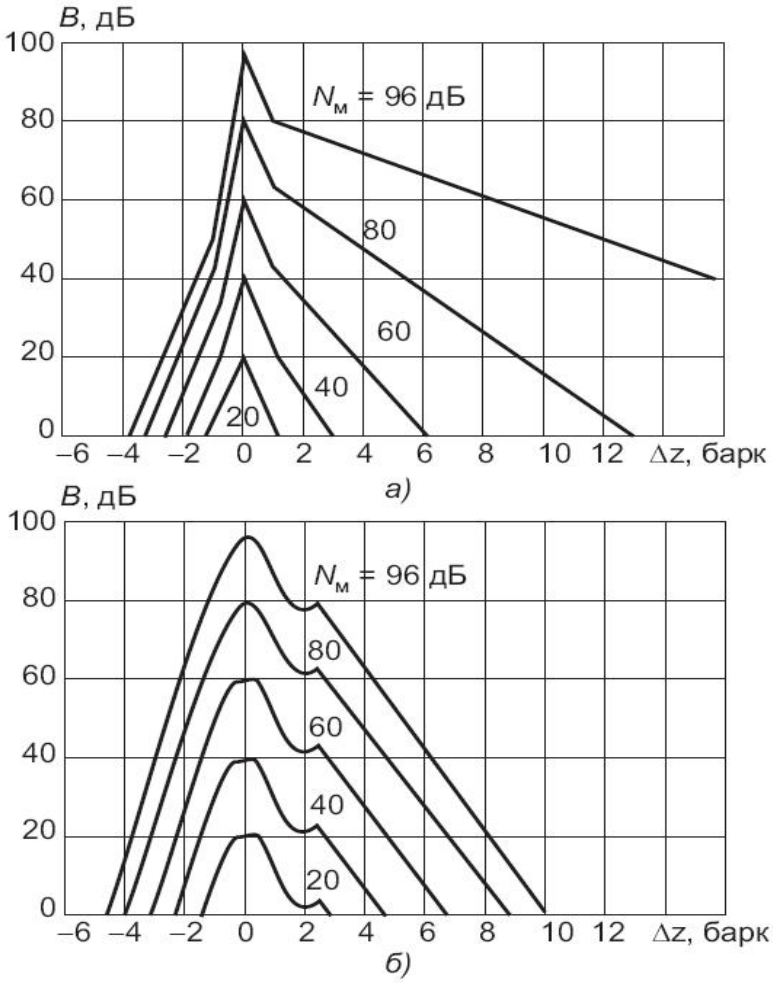


Рис. 1.37. Аппроксимации кривых маскировки (развертывающих функций) для различных уровней маскирующего тона: *а* – психоакустическая модель 1 стандартов *MPEG ISO*/*IEC* 11172-3 и 13818-3; *б* – психоакустическая модель 2 стандартов *MPEG ISO*/*IEC* 11172-3, 13818-3, 14496-3

функций обладают незначительными взаимными расхождениями лишь для среднего уровня маскирующей компоненты. Наибольшие расхождения кривых наблюдаются для малого и высокого уровней маскирующих компонент. Поэтому для принятия решения о том, какая аппроксимация является наиболее адекватной слуховому восприятию необходимо провести тестовые прослушивания для каждого такого случая. Будем при выборе аппроксимирующего выражения для развертывающей функции руководствоваться следующими сведениями о свойствах кривых маскировки, полученных экспериментальным путем:

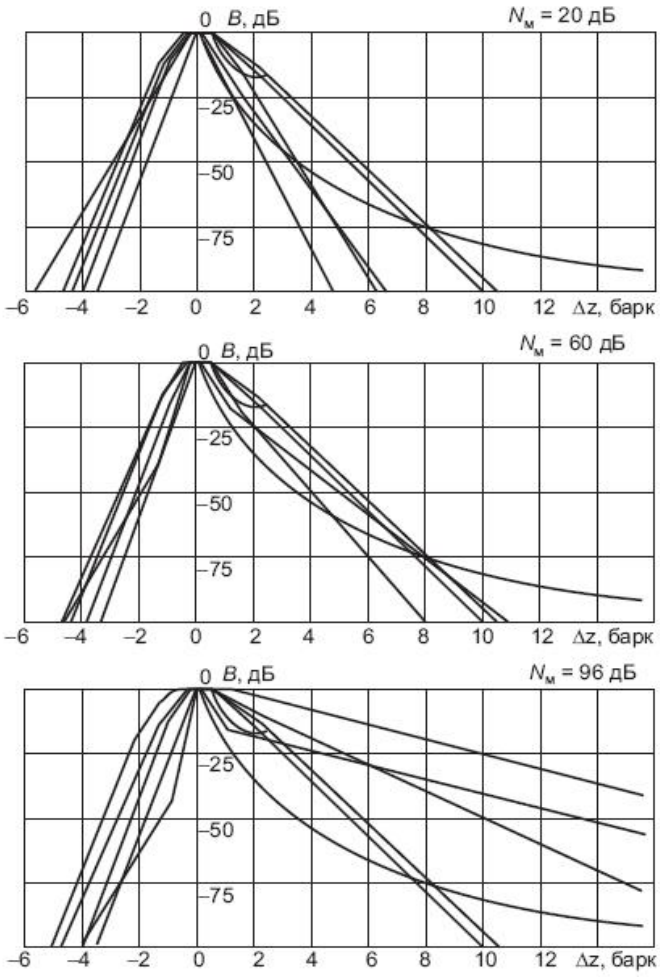


Рис. 1.38.Нормированные развертывающие функции для уровня маскирующей компоненты $N_м$ 20, 60 и 96 дБ

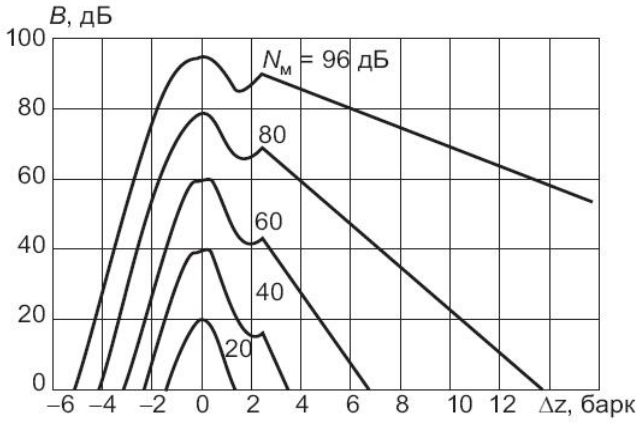


Рис. 1.39. Семейство кривых маскировки для различных уровней энергии сигнала в *b*-ой полосе психоакустического анализа, наилучшим образом соответствующее экспериментальным данным

-крутизна переднего фланца кривой относительного порога слышимости является инвариантной по отношению к уровню маскирующей компоненты, [1.27];

-крутизна заднего фланца кривой относительного порога слышимости зависит от уровня маскирующей компоненты, [1.4];

-должно быть учтено локальное нелинейное изменение относительного порога слышимости вблизи маскирующей компоненты со стороны заднего фланца, [1.4].

Анализ предлагаемых аппроксимаций развертывающих функций показал, что указанными свойствами обладает кривая, имеющая следующую аналитическую запись:

$$B(\Delta z) = A_0 + A_1 + \left(\frac{D_1 + D_2}{2}\right)\cdot(\Delta z + A_2) - \left(\frac{D_1 - D_2}{2}\right)\cdot\sqrt{\gamma + (\Delta z + A_2)},$$

$$A_0 = 8\cdot\min\left((\Delta z - 0.5)^2 - 2(\Delta z - 0.5), 0\right),\ A_1 = \sqrt{\gamma\cdot D_1\cdot|D_2|},$$

$$A_2 = \frac{D_1 + D_2}{2}\cdot\sqrt{\frac{\gamma}{D_1\cdot|D_2|}},\ D_1 = 27\ \text{дБ/барк},\ D_2 = -(22 - 0{,}2\cdot N_M)\ \text{, дБ/барк},$$

$$\gamma = 1.$$

Графически семейство данных кривых маскировки для различных уровней энергии маскирующего сигнала в *b*-ой полосе психоакустического анализа представлено на рис. 1.39.

Индекс тональности. Выше рассмотрены предельные случаи, когда сигналы представляют собой шум или тон. В реальных условиях чаще всего встречаются промежуточные состояния. Количественной мерой схожести реального сигнала с тональным или шумовым может служить коэффициент (индекс) тональности α. Для чистого тона $\alpha = 1$, для шума $\alpha = 0$.

С учетом индекса тональности общую формулу для коэффициента маскировки можно представить в виде [1.10]:

$$\Delta N_{\mathrm{M},i} \approx -\,[\alpha(14{,}5+i)+(1{,}0-\alpha)\cdot 5{,}5]\ \text{, дБ.}$$

Значение коэффициента маскировки $\Delta N_{\mathrm{M},i}$ меняется от - 14,5 дБ для первой частотной группы (i = 1) до - 38,5 дБ для частотной группы i = 24. Для шумоподобного сигнала (тон маскируется шумом) α = 0 и величина коэффициента маскировки ΔN_{M2} в первом приближении равна минус 5,5 дБ и не зависит от частоты тонального сигнала.

Если учесть зависимость коэффициента маскировки от частоты, то получим более точный результат, определяемый выражением [1.10]:

$$\Delta N_{M,i} \approx -[\alpha(i)\cdot(14{,}5+i)+(1-\alpha(i))\cdot|\Delta N_{M2}(i)|], \text{ дБ},$$

где i – номер частотной группы слуха (i = 1,2,...,24), $\alpha(i)$ – коэффициент тональности маскирующего сигнала. Значение коэффициента тональности в стандартах *MPEG* вычисляется для каждой частотной группы i.

Коэффициент Хаоса. (*Chaosmass – measure of chaos*, K_X). Он оценивает степень близости маскирующего сигнала к тональному или шумоподобному сигналам. Величина коэффициента Хаоса для реальных сигналов изменяется от 0 до 0,5. При этом сигнал считается тональным, если значение K_X лежит в пределах от 0 до 0,05. Между значениями α и K_X существует связь, для ее оценки предложены два выражения [1.10]:

$$\alpha \approx -0{,}299-0{,}43\cdot\log_e(K_X) \text{ и } \alpha \approx -0{,}2686-0{,}2\cdot\log_e(0{,}5\cdot K_X)$$

Зависимости, соответствующие данным аппроксимирующим функциям, представлены на рис. 1.40 (кривые 1 по [1.26] и 2 по [1.10]). По оси абс-

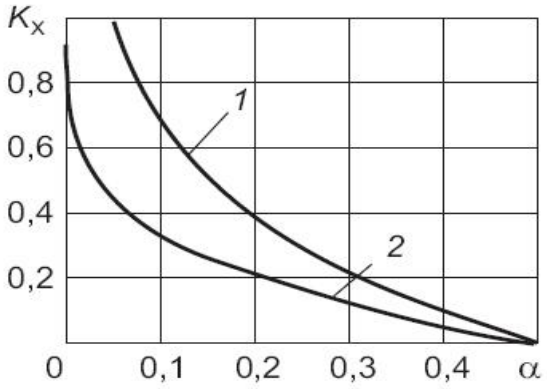


Рис. 1.40. Значения коэффициента хаоса по данным [1.26], кривая 1 и [1.10], кривая 2

цисс отложены значения коэффициента Хаоса K_X, по оси ординат - значения коэффициента тональности α. Видно, что зависимости, полученные разными авторами, имеют существенные расхождения. Процедура расчета K_X достаточно сложна [1.10]. Потому в психоакустических моделях систем кодирования с компрессией цифровых аудиоданных (см. например, стандарты *ISO/IEC* 11172-3 и 13818-3) при оценке субполосных составляющих введены существенные упрощения.

1.7. Маскировка во временной области

Кроме одновременной маскировки, слуху присуща и маскировка во времени. В этом случае маскирующее воздействие ограниченного во

времени выброса звукового сигнала выходит за пределы моментов его физического возникновения и прекращения. Другими словами, относительные пороги слышимости для замаскированных звуков увеличены не только в течение, но и до и после возникновения выброса маскирующего сигнала (рис. 1.41, где по оси ординат отложены уровни сигнала, дБ, а по оси абсцисс – текущее время, мс).

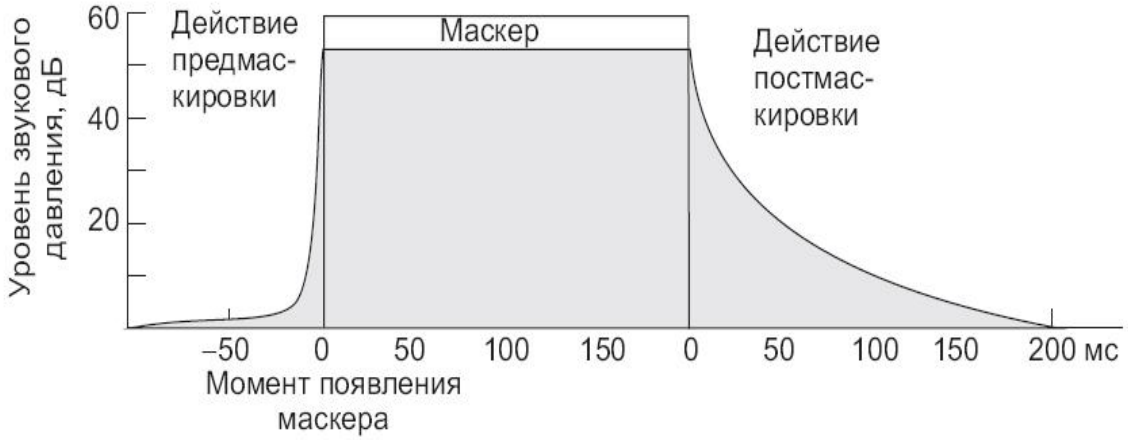


Рис. 1.41. Схематичное представление эффектов временной маскировки [1.6]: верхняя часть затемненной области показывает значение относительного порога маскировки для сигнала, ограниченного во времени; уровень маскирующего сигнала составляет 60 дБ

В соответствии с данными рис. 1.41 различают предмаскировку и постмаскировку (послемаскировку).

Предмаскировка (*backward masking*) – это увеличение порога слышимости при восприятии звукового сигнала на некотором временном интервале непосредственно перед выбросом в его временной функции, данное явление проявляет себя на временном отрезке длиной 8…10 мс.

Постмаскировка (*forward masking*) – это увеличение порога слышимости при восприятии звукового сигнала на некотором временном интервале непосредственно сразу после выброса в его временной функции (рис. 1.42); это явление проявляет себя на временном интервале до 200…250 мс, считая с момента окончания действия выброса.

Предмаскировка проявляется на значительно более коротком временном интервале, он обычно не превышает 8…10 мс. Длительность предмаскировки в очень сильной степени зависит от особенностей индивидуума. Чаще всего именно по этим двум причинам явление предмаскировки в алгоритмах компрессии не учитывается.

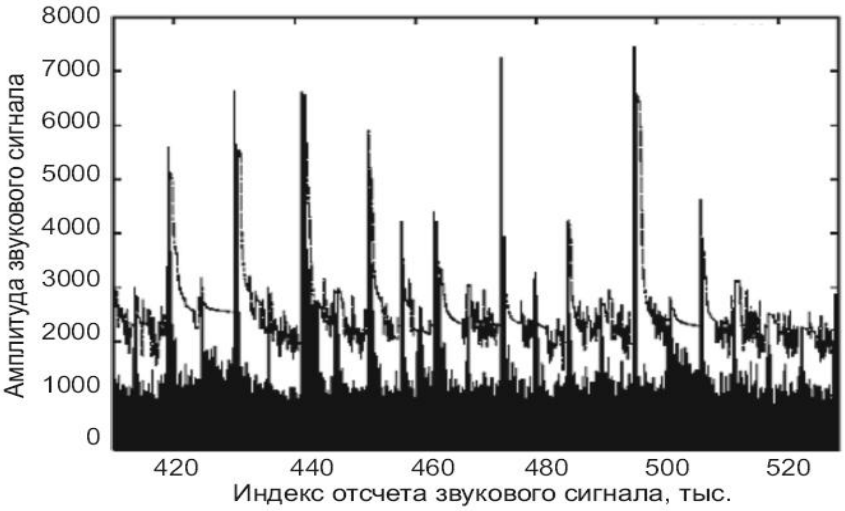


Рис. 1.42. Пример отрывка временной функции звукового сигнала: пунктир – значения относительного порога слышимости с учетом постмаскировки [1.28]

Моделирование постмаскировки

Из публикаций известно, что феномен постмаскировки обладает следующими свойствами:

-относительные пороги слышимости после окончания действия выброса сигнала (маскера) плавно спадают с увеличением текущего времени от своего максимального значения, вызванного воздействием выброса на слуховую систему человека;

-узкополосный шум провоцирует более высокое начальное значение порога постмаскировки, чем тональный сигнал того же уровня;

-скорость спада кривой относительного порога слышимости зависит как от уровня, так и от длительности выброса маскирующего сигнала;

-наклон кривой порога слышимости изменяется с увеличением временного сдвига между маскирующим и маскируемым сигналами;

- быстрый спад маскирующего воздействия характерен для небольшой задержки во времени между маскирующим и маскируемым сигналами; более медленный спад отмечен для больших временных задержек;

-крутизна спада порога слышимости, вызванного постмаскировкой, на начальном этапе обладает частотной зависимостью;

-постмаскировка зависит от вида маскирующего сигнала: тональные маскирующие сигналы после прекращения их действия оказывают более длительное воздействие, чем узкополосные шумоподобные сигналы;

-явление постмаскировки проявляется и в частотной области.

Физиологическая причина возникновения временной постмаскировки до сих пор точно не установлена. В [1.30, *W.Jesteadt*] высказано

предположение о том, что мембрана слухового анализатора человека формирует значение порога постмаскировки для коротких (до 10…20 мс) временных сдвигов непосредственно после прекращения действия маскирующего сигнала. За более длительный эффект отвечает уже нейронная обработка в высших отделах слухового анализатора. В [1.6, *E.Zwicker*, *H.Fastl*] также говорится об интегрирующей способности слуха, ведущей к эффектам временной маскировки.

Наиболее полное исследование психоакустических закономерностей временной маскировки выполнено Х.Фастлом [1.31,…,1.33]. Оно охватывает исследование частотных свойств временной маскировки, взаимодействие (взаимосвязь) эффектов временной и одновременной маскировки, а также механизмы суммирования относительных порогов маскировки во временной области. В частности, он представил трехмерные схематические модели изменения относительных порогов маскировки (рис. 1.43), создаваемые ограниченным по времени воздействия маскирующим сигналом.

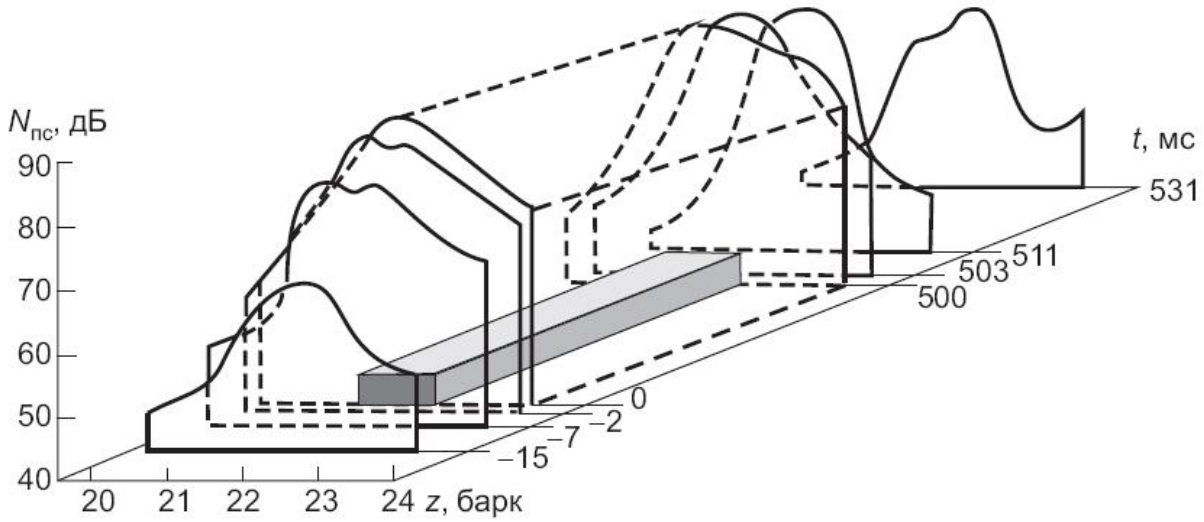


Рис. 1.43. Изменения порога слышимости при воздействии узкополосного шумового маскирующего сигнала со средней частотой 8,5 кГц, уровнем 70 дБ и длительностью 500 мс [1.32]

Все данные, относящиеся к частотной области, для большего удобства исследователей, разрабатывающих алгоритмы компрессии цифровых аудиоданных, представлены в шкале высот тона z, оцениваемой в барках. По второй горизонтальной оси на рис. 1.43 отложено текущее время t, мс, по вертикальной оси – уровень порога слышимости $N_{ПС}$, дБ. Здесь упрощенно показано развитие всех трех процессов: предмаскировки, маскировки при одновременном воздействии маскирующего и маскируемого сигналов и постмаскировки, показывающей изменения порога слышимости уже после окончания воздействия маскирующего сигнала. В результате этого и ряда других экспериментальных исследований стало возможным получение более точных аналитических

зависимостей, учитывающих влияние уровня, частоты, и длительности маскирующего сигнала на изменение порога слышимости после окончания звукового воздействия. Полученные данные многократно проверялись многими исследователями, затем на их основе создавались математические модели, описывающие эти закономерности, что делало возможным использование полученных данных на практике, например, при разработке эффективных алгоритмов компрессии цифровых аудиоданных. Но отсутствие единого подхода к экспериментам и условиям их проведения привело к тому, что количественные результаты, полученные разными авторами, часто существенно различались. Именно это в первую очередь осложняет построение достоверных математических моделей, описывающих с требуемой для практики точностью, эффекты временной маскировки.

Можно выделить три основных модели, наиболее полно и достоверно аппроксимирующие психоакустические закономерности постмаскировки и получившие наибольшее распространение:

-аппроксимация данного явления логарифмической функцией времени, основанная на результатах ранних исследований Пломба [1.34, *R.Plomb*] и Штайна [1.35, *H.J.Stein*];

-аппроксимация постмаскировки суммой экспоненциальных зависимостей с различной скоростью спада таких кривых;

-аппроксимации, основанные на экспериментальных данных Э.Цвикера [1.4; 1.6], учитывающие зависимость порога постмаскировки от длительности маскирующего импульса.

Рассмотрим эти модели более подробно. Исследования Р. Пломба [1.34] показали, что зависимости постмаскировки достаточно хорошо аппроксимируются прямой линией, если значение текущего времени взято в логарифмическом масштабе. При этом длительность постмаскировки почти не зависит от уровня маскирующего сигнала, что спорно. Развивая эту идею, *W.Jesteadt*, *S.P. Bacon* и *J.R.Lehman* [1.30] предложили аппроксимирующую функцию порога постмаскировки следующего вида:

$$D(t, N_M) = a\left(b - \log t\right)\left(N_{\text{М}} - c\right), \text{ дБ}$$

где $N_{\text{М}}$ – уровень маскирующего сигнала, дБ; t – время, прошедшее с момента его выключения, мс; *a, b, c* – константы. Изменяя величины *a,b, c,* они добились хорошего приближения кривой спада маскирующего воздействия для различных частот к результатам проводимых ими экспериментов. При этом они установили, что на низких частотах (в полосе 125…250 Гц) эффект постмаскировки выражен значительно более сильно, что приводит к более значительному повышению порогов маскировки после прекращения воздействия маскирующего сигнала, чем на частотах

выше 500 Гц. На рис. 1.44 приведены четыре группы кривых постмаскировки – соответственно для частот 125, 250, 1000 и 4000 Гц. Параметром представленных кривых является уровень $N_М$ тонального маскирующего сигнала, равный 30, 50 или 70 дБ. За начало отсчета текущего времени взят момент выключения маскирующего тона.

Модель, предложенная в [1.30, *Jesteadt W.*, *Bacon S. P.*, *Lehman J. R.*], развита и дополнена экспериментальными данными [1.36, *Moore B.C.J.*, *Glasberg B. R.*]. В [1.36] утверждается, что, несмотря на хорошую в целом аппроксимацию кривых постмаскировки такой зависимостью, скорость спада этих кривых для малых значений текущего времени (менее 20 мс) может быть для тональных маскирующих сигналов заметно выше значений, представленных в [1.30].

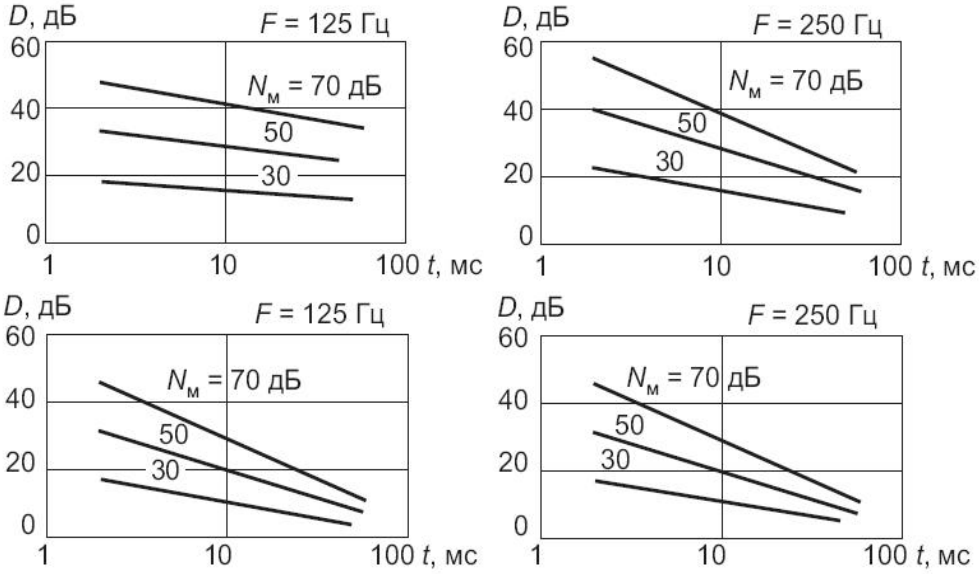


Рис. 1.44. Аппроксимация порогов постмаскировки для тонального маскирующего сигнала с частотами 125, 250, 1000 и 4000 Гц и различными уровнями $N_М$ [1.30]

Фундаментальную работу по исследованию влияния длительности маскирующего сигнала на изменение порога посмаскировки опубликовали Э. Цвикер и Х. Фастл [1.37]. Ими были проведены измерения относительных порогов слышимости для широкополосного маскирующего белого шума. При этом длительность его воздействия $T_М$ составляла 5, 10, 30 и 200 мс (рис. 1.45). Набольшее изменение порога маскировки было отмечено при длительности шумового воздействия, лежащей в пределах от 5 до 30 мс. Максимальное повышение порогов слышимости составляет при этом 16 дБ. Изменение длительности маскирующего сигнала от 30 до 200 мс приводит к менее значительному влиянию на кривые постмаскировки, что говорит о достижении предельного максимума этой зависимости при длительности звукового воздействия около 200 мс. Для маскирующих шумовых сигналов длительностью воздействия более 200 мс влияния их

длительности на крутизну кривых постмаскировки уже не отмечалось. Заметим также, что для тональных маскирующих сигналов не отмечалось столь существенного влияния длительности маскирующего сигнала на крутизну спада кривых постмаскировки.

Исследования Э. Цвикера и Х. Фастла послужили далее основой для нескольких математических моделей временной постмаскировки. Аппроксимирующая функция временной посмаскировки, учитывающая влияние длительности маскирующего воздействия, предложена Капустом [1.10, *Kapust R.*]. В своих исследованиях он, в частности, основывался на экспериментальных данных Э. Цвикера [1.4]:

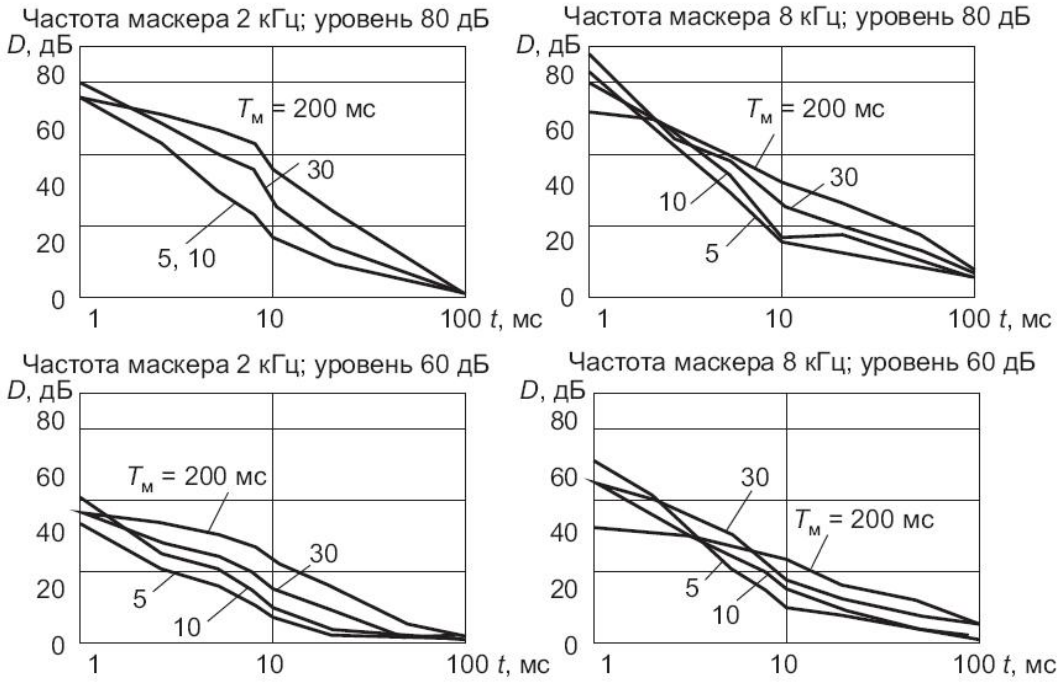


Рис. 1.45. Пороги постмаскировки для различных частот, уровней и длительностей тонального маскирующего сигнала [1.37]

$$D(t,T) = 1,0 - \left(\frac{1}{1,35}\right) \cdot arctg\left(\frac{t}{13,2 \cdot (T)^{-0,25}}\right),$$

где T – длительность маскирующего сигнала, мс; t – время, прошедшее с момента его выключения, мс. Сравнивая эту модель с наборами экспериментальных данных [1.4], можно констатировать достаточно точную аппроксимацию зависимостей постмаскировки (рис. 1.46). Параметром каж-

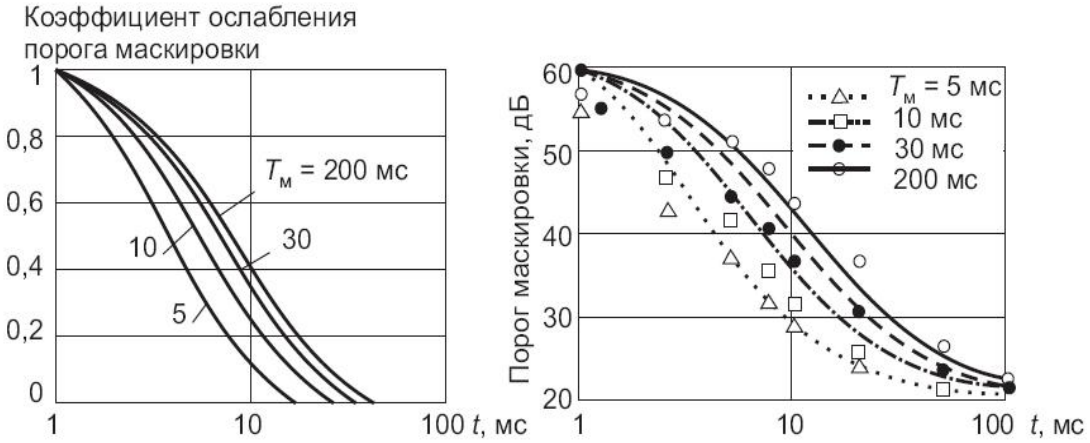


Рис. 1.46. Пороги постмаскировки для различных частот, уровней и длительностей тонального маскирующего сигнала [1.6, 1.10]

дой кривой является длительность маскирующего сигнала T = 5, 10, 30 или 200 мс. Видно, что постмаскировка проявляется на интервале времени равном 100...200 мс. Различными значками здесь отмечены уточненные значения порогов маскировки [1.4] для узкополосного маскирующего шума с уровнем 60 дБ. Постоянные аппроксимирующей функции подобраны для достижения наилучшего совпадения.

Однако вышеприведенное выражение требует уточнения с целью учета и других факторов, существенно влияющих на кривые постмаскировки, таких, как частота и уровень маскирующего воздействия.

Известны также и другие аппроксимации кривых постмаскировки (табл.1.4).

Таблица 1.4

Аппроксимирующие функции для расчета постмаскировки

Номер литературного источника	Аппроксимирующее выражение
[1.30]	$D(t,N_M)=a(b-\log t)(N_{\text{М}}-c)$, где $N_{\text{М}}$ – уровень маскирующего сигнала, дБ; t – время, прошедшее с момента его выключения, мс; a, b, c – константы
[1.10], [1.4]	$D(t,T)=1,0-\left(\frac{1}{1,35}\right)\cdot arctg\left(\frac{t}{13,2\cdot(T)^{-0,25}}\right)$, где T – длительность маскирующего сигнала, мс; t – время, прошедшее с момента его выключения, мс
[1.45]	$D(t)=55,5955\cdot 10^{(-0.0163\cdot t)}$ $D(t)=58,4039\cdot 10^{(-0.0059\cdot t)}$, где t- задержка в мс; длительность маскирующего сигнала соответственно 5 и 200 мс

Окончание табл. 1.4

Номер литературного источника	Аппроксимирующее выражение
[1.40],[1.46]	$D(t,N_M)=k\cdot 10\log\left(N_{\text{М}}^{\left(e^{-t/\lambda_1}+e^{-t/\lambda_2}\right)}+N_0\right)$, дБ, где λ_1 и λ_2 – константы, определяющие крутизну спада кривых постмаскировки; $N_{\text{М}}$ – уровень маскирующего воздействия; N_0 – абсолютный порог слышимости; k – корректирующий множитель; t – текущее время, прошедшее с момента выключения маскирующего сигнала
[1.41]	$N_{\text{ПМ}}=k\left(N_{\text{М}}+m\right)e^{-t/T}$, где $N_{\text{ПМ}}$ – порог постмаскировки, дБ; $N_{\text{М}}$ – уровень маскирующего воздействия, дБ; k – константа, корректирующая крутизну спада кривой постмаскировки; m– константа, определяющая чувствительность, или индекс маскировки
[1.42]	$D(t)=(1-w)\cdot\left(1+\frac{2t}{\tau_p}\right)\cdot\exp\left(-\frac{2t}{\tau_p}\right)+w\cdot\left(1+\frac{2t}{\tau_S}\right)\cdot\exp\left(-\frac{2t}{\tau_S}\right)$ где τ_P-константа, определяющая крутизну спада для коротких задержек маскируемого и маскирующего сигналов (менее 30 мс); τ_S- константа, определяющая крутизну спада кривой постмаскировки на остальном временном интервале; w - весовой коэффициент; t - время, прошедшее с момента выключения маскирующего сигнала
[1.35], [1.43]	$D(t)=(1-w)\cdot\exp\left(-\frac{t}{\tau_1}\right)+w\cdot\exp\left(-\frac{t}{\tau_2}\right)$, где τ_1, τ_2 – константы (постоянные времени), определяющие крутизну спада кривой постмаскировки; w - весовой коэффициент; t – время, прошедшее с момента выключения маскирующего сигнала

Зависимость постмаскировки от частоты

Зависимость формы кривой постмаскировки от частоты и уровня маскирующего сигнала удалось хорошо учесть, изменяя постоянные времени и весовые коэффициенты для двух экспоненциальных составляющих выражения [1.35, *Stein H. J.*], табл. 1.4, нижняя строка. В [1.43, *Plack C.J.*, *Moore B.C.J.*] приведены наборы параметров для описания данного явления (рис. 1.47).

Таблица 1.5

Параметры аппроксимирующей функции

Частота маскрующего сигнала, Гц	300			900			2700			8100		
Уровень маскрующего сигнала, дБ	64	54	44	52	42	32	40	30	20	49	39	29
τ_1, мс	5.5	4.9	4.5	2.5	3.5	4.0	2.4	2.6	2.9	2.1	2.9	2.9
τ_2, мс	19	36	27	13	32	28	14	14	17	14	18	17
w, дБ	-69	-56	-50	-51	-51	-39	-44	-39	-34	-46	-40	-33

Здесь представлены зависимости порога постмаскировки, полученные для четырех тональных маскирующих сигналов разных уровней с частотами 300, 900, 2700 и 8100 Гц. При этом их уровни для каждой из частот выбраны с учетом получения эквивалентных уровней возбуждения. По горизонтальной оси отложено текущее время, мс. При этом значения параметров аппроксимирующей функции (табл. 1.4, нижняя строка) представлены ниже (табл. 1.5).

В [1.44, *Plack C. J., Oxenham*] данная функция (табл.1.4, нижняя строка) упрощена и получено вполне приемлемое совпадение с экспериментальными данными (рис. 1.48):

$$D(t) = 0{,}975\exp\left(\frac{t}{4}\right) + 0{,}025\exp\left(\frac{t}{29}\right),$$

где t – текущее время. Очевидно, что такая аппроксимация постмаскировки не учитывает ее зависимости ни от длительности маскирующего сигнала, ни от его частоты. Установлено, что порог маскировки не увеличивается

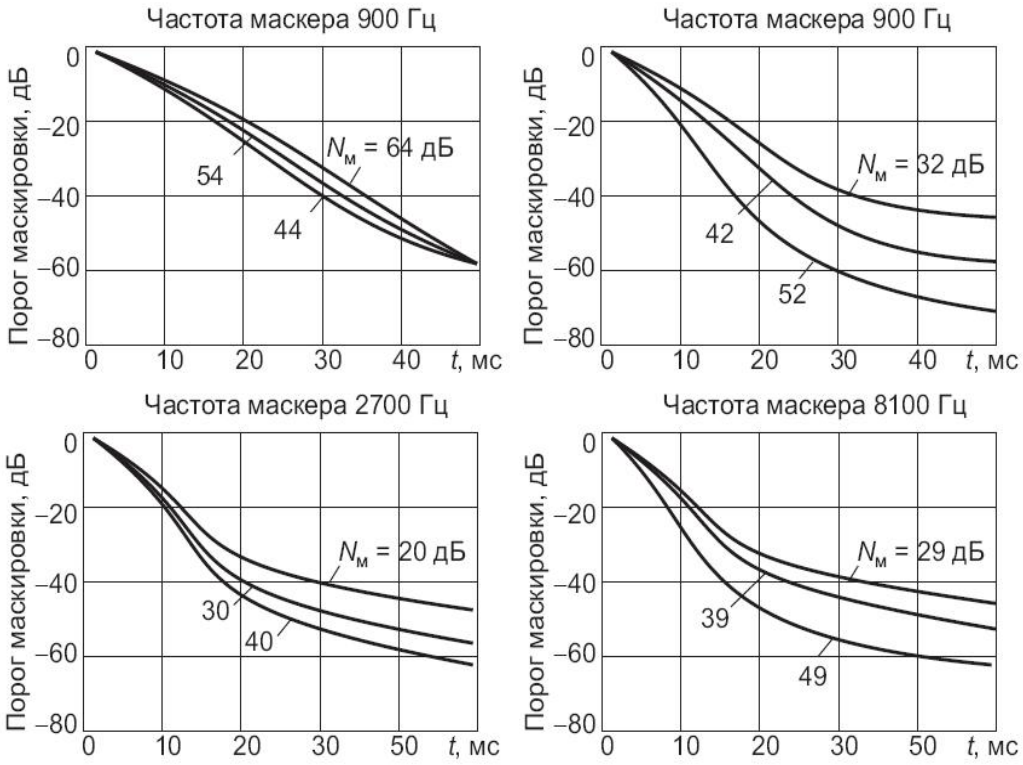


Рис. 1.47. Аппроксимация явления постмаскировки для различных частот и уровней маскирующего сигнала [1.43]

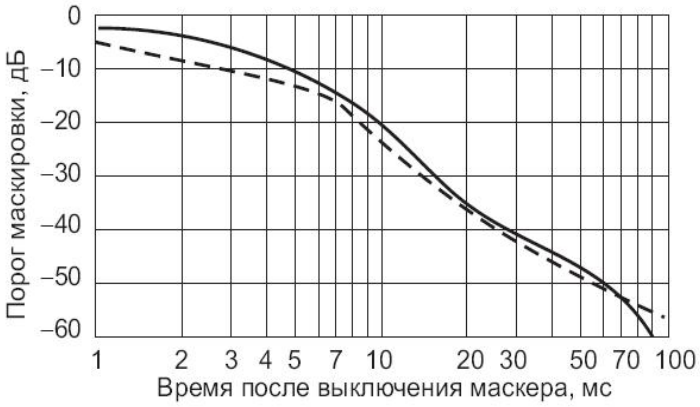


Рис. 1.48.Кривые постмаскировки: верхняя линия-аппроксимация [1.44]: пунктир – экспериментальные данные для шумового маскирующего сигнала с уровнем 80 дБ и длительностью воздействия 200 мс

линейно с увеличение уровня маскирующего сигнала, как при одновременной маскировке, а нарастает значительно медленнее. Для упрощенного учета такой нелинейной зависимости кривой постмаскировки от уровня маскирующего сигнала в [1.44] было предложено (см. выше) предварительно обработать входные значения, поступающие на вход этой функции, компримирующей функцией с двумя участками (рис. 1.49):

$N_{\text{вых}} = 0,78N_{\text{вх}}$ для маскеров с уровнями $N_{\text{вх}}$< =35 дБ,

$N_{\text{вых}} = 0,16N_{\text{вх}} + 21,7$ для маскеров с уровнями $N_{\text{вх}}$ > 35 дБ,

где $N_{\text{вх}}$ и $N_{\text{вых}}$ – исходный и преобразованный (взвешенный) уровни, дБ. Константа 21,7 обеспечивает непрерывность величин $N_{\text{вых}}$ при переходе через пороговое значение, равное 35дБ. На рис. 1.49 по оси абсцисс отложены фактические входные уровни $N_{\text{вх}}$ маскирующего сигнала, дБ, а по оси ординат их преобразованные значения $N_{\text{вых}}$, также в дБ.

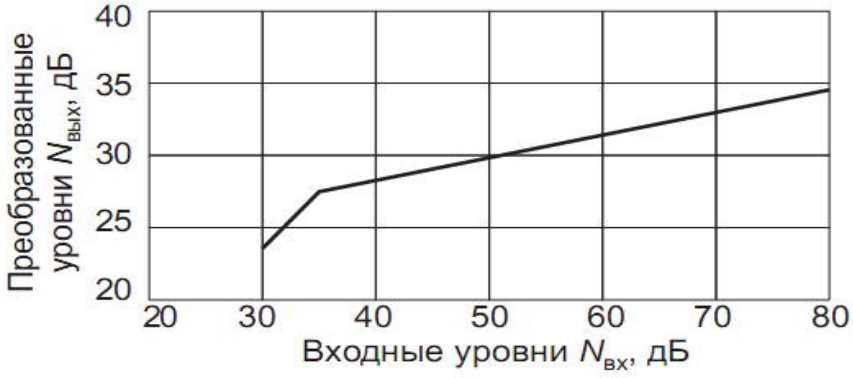


Рис. 1.49. Компрессия постмаскировки по Плаку-Оксенхаму [1.44]

В рекомендации *ITU-R BS*1387 [1.47] описывается метод объективной оценки качества кодирования аудиосигналов, основанный на психоакустическом анализе. При этом для учета свойств временной маскировки используются сглаживающие фильтры первого порядка. Как известно, они обладают экспоненциальным спадом импульсной характеристики. В то же время многие исследователи [1.6] прямо указывают, что постмаскировка не может моделироваться строго экспоненциальным спадом относительного порога слышимости.

Предлагаемый в [1.47] банк сглаживающих фильтров является частотно-зависимым. Его постоянные времени заданы выражением:

$$t_i = t_{\min} + (100/F_{cp}) \cdot t_{100}, \quad \text{причем} \quad \left| \begin{array}{l} t_{\min} = 0,008 \\ t_{100} = 0,030 \end{array} \right.$$

где $F_{\text{ср}}$ – центральная частота фильтра в Гц. В качестве примера на рис. 1.50 показаны графики спада порогов постмаскировки для нескольких

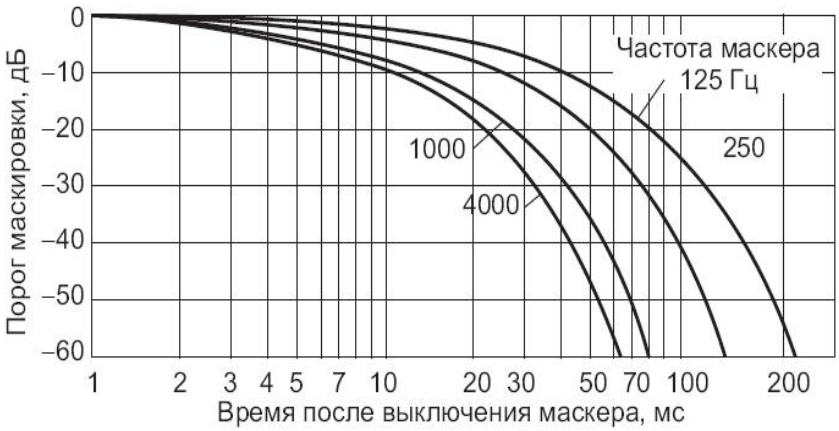


Рис. 1.50. Кривые порога постмаскировки по ITU-R BS1387

выбранных частот. Установлено также, что постмаскировка обладает несколько большей частотной селективностью, чем одновременная маскировка, особенно в направлении возрастания частоты вверх от ее значения для маскирующего сигнала, что приводит к более быстрому спаду маскирующего воздействия в сторону возрастания частоты, чем при одновременной маскировке [1.49]. Так, Э. Цвикер и Х. Фастл, исследуя спад относительного порога постмаскировки тонального маскирующего сигнала в зависимости от его длительности [1.6], установили, что крутизна спада относительных порогов одновременной маскировки в общем сохраняется и после прекращения действия маскирующего сигнала, т.е. при постмаскировке. При этом скорость спада порога постмаскировки (рис. 1.51) с увеличением задержки от маскирующего сигнала в сторону

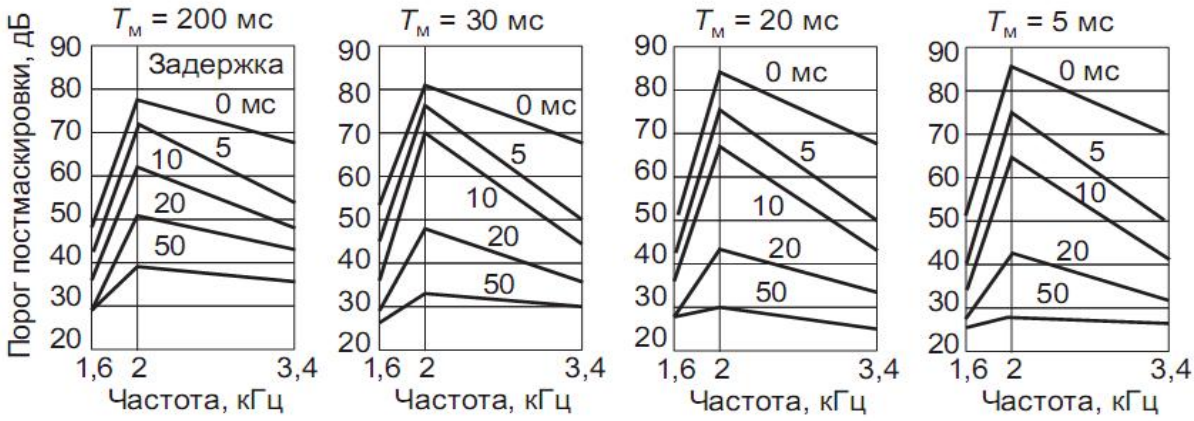


Рис. 1.51. Зависимости распространения постмаскировки в частотной области от времени, прошедшего после выключения маскирующего сигнала. В качестве маскирующего сигнала взят узкополосный шум с центральной частотой 8,5 кГц, уровнем 70 дБ, длительность 500 мс с шириной в одну критическую полосу слуха (по данным Х.Фастла)

уменьшения частоты сохранялась постоянной, а в сторону увеличения частоты несколько увеличивалась. Это указывало на большую частотную избирательность постмаскировки, по сравнению с одновременной маскировкой, что и подтвердили дальнейшие исследования, например [1.50, *Moore B.C.*].

И последнее. Наложение порогов постмаскировки и предмаскировки для последовательно следующих во времени маскирующих сигналов ведет к существенному дополнительному повышению порогов слышимости на 10…15 дБ по сравнению с простым их сложением. Этот эффект обусловлен интегрирующей способностью слуха [1.31…1.33].

В отличие от одновременной маскировки, где наличие нескольких маскирующих сигналов всегда ведет к увеличению маскирующего воздействия, при постмаскировке возможно и уменьшение порога слышимости из-за присутствия дополнительного маскирующего сигнала, например тона, на 3 дБ и более [1.51, *Houtgast. T.*]. Эффект уменьшения порога постмаскировки из-за присутствия другого маскирующего сигнала, удаленного по частоте, получил название «*laterale suppression*».

Реализация моделей учета постмаскировки

Для учета явления постмаскировки может быть использована *RC*-цепь [1.38], рис. 1.52.

Одним из свойств постмаскировки является зависимость величины относительного порога слышимости от амплитуды и длительности выброса временной функции сигнала. Поэтому для анализа работы RC-цепи рассмотрим зависимость изменения во времени напряжения на выходе такой

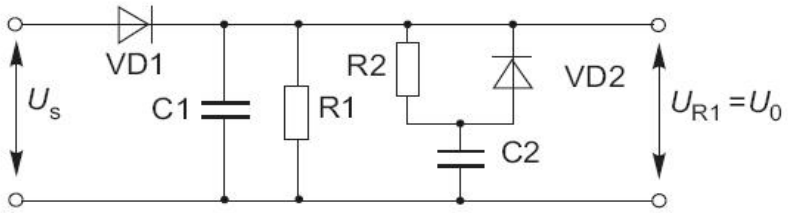


Рис. 1.52. Принципиальная схема RC – цепи, моделирующей явление постмаскировки (R1=35 кОм, C1=0.7 мкФ, R2=20 кОм, C2=1 мкФ)

цепи при воздействии на ее вход импульсной посылки постоянного напряжения определенной амплитуды и длительности ΔT.

Пусть на вход *RC*-цепи воздействует напряжение, прямоугольной формы с амплитудой *U*, которое заряжает конденсаторы C1 и C2. Заряд конденсатора C1 осуществляется через диод VD*1* (сопротивление открытого *p-n*-перехода которого составляет $R_{ОП}$), а конденсатор C2 заряжается через диод VD*1* и резистор R2.

Величина напряжения, до которого заряжаются конденсаторы C1 и C2, зависит к моменту окончания действия импульсной посылки от ее длительности ΔT, а также от значений постоянных времени цепей заряда C1 и C2. При малом значении ΔT величина напряжения на конденсаторе C1 значительно превышает напряжение на конденсаторе C2, при длительности импульсной посылки превышающей 300 мс конденсаторы C1 и C2 практически заряжены полностью и напряжения на них равны.

По окончании действия на вход *RC*-цепи (рис. 1.52) импульсной посылки постоянного напряжения, конденсаторы C1 и C2 начинают разряжаться, причем конденсатор *C1* разряжается непосредственно через резистор R1, а конденсатор C2 – либо через резисторы R2 и R1, либо через диод VD2 и резистор R1. Если падение напряжения на резисторе R1 превышает величину напряжения на конденсаторе C2, то диод VD2 закрыт и разряд осуществляется только через резистор R2, по мере разряда конденсатора C1 напряжение на резисторе R1 уменьшается и, как только оно становится меньше напряжения на конденсаторе C2, открывается диод VD2 и конденсатор C2 начинает разряжаться через открытый диод.

Кривые, показывающие изменение напряжения на выходе этой цепи при воздействии на ее вход импульсной посылки постоянного напряжения длительностью ΔT, представлены на рис. 1.53. По оси ординат отложены нормированные по отношению к максимальному значению величины напряжения на выходе цепи. По оси абсцисс – текущее время, мс. Параметром каждой зависимости является длительность импульсной посылки постоянного напряжения ΔT, мс, причем ΔT = 5, 10, 50, 100, 200, 300 мс.

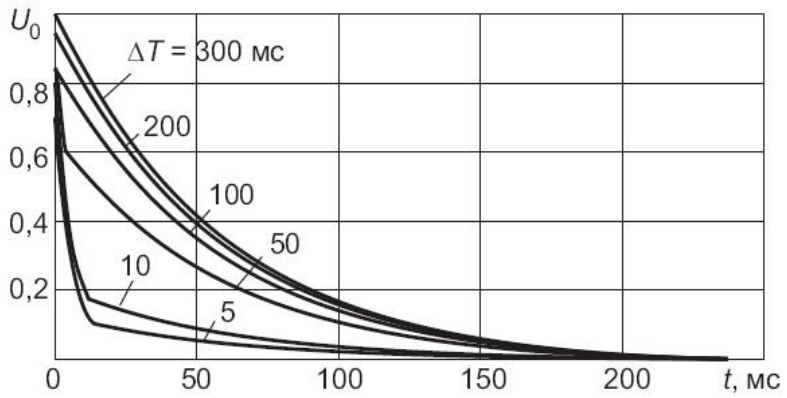


Рис. 1.53. Семейство зависимостей изменения во времени напряжения на выходе RC - цепи при воздействии на ее вход импульсной посылки постоянного напряжения постоянной амплитуды, но разной длительности

Анализируя работу данной *RC*-цепи при различных длительностях ΔT воздействия на ее вход импульсной посылки постоянного напряжения (рис. 1.53), можно сделать следующие выводы:

-при малых длительностях воздействия скорость спада напряжения на выходе определяется преимущественно временем разряда конденсатора *C1* (кривые, соответствующие значениям ΔT равным 5 и 10 мс). При больших длительностях скорость спада выходного напряжения определяется уже временем разряда параллельно подключенных конденсаторов С1 и С2 (кривые, соответствующие значения ΔT равным 100, 200 и 300 мс);

-с ростом длительности воздействия ΔT, скорость спада выходного напряжения уменьшается. Это обусловливается соотношением энергий, запасенных в конденсаторах С1 и С2;

-независимо от длительности и амплитуды импульсной посылки постоянного напряжения функция выходное напряжение асимптотически стремится к нулю через интервал времени $\approx$ 300 мс;

-изменение скорости спада напряжении на выходе *RC*-цепи наблюдается только для длительности воздействия импульсной посылки не превышающей $\approx$ 200 мс.

При моделировании явления постмаскировки данной *RC*-цепью не требуется дополнительного знания амплитуды и длительности выброса временной функции сигнала, т.к. спад кривых постмаскировки в этом случае зависит только от запасенной в этой цепи энергии. Напомним, что явление постмаскировки может быть объяснено инерцией основной мембраны человеческого уха. Рассматриваемая *RC*-цепь позволяет моделировать это ее свойство.

При моделировании явления постмаскировки необходимо также иметь в виду, что скорость спада кривых, учитывающих это явление, должна зависеть и от частоты воздействующего колебания. Об этом свидетельствуют экспериментальные данные, приведенные в [1.30], где было показано, что скорость спада порога слышимости, вызванного явлением постмаскировки,

уменьшается с увеличением амплитуды и возрастает с ростом частоты звукового воздействия. Наличие частотной зависимости кривых постмаскировки может быть учтено дополнительно с помощью коэффициента τ. Он определяет время, за которое порог постмаскировки уменьшается в е ≈ 2,72 раз от своего максимального значения.

Аналитическое выражение, учитывающее зависимость скорости спада кривой постмаскировки от частоты F, имеет следующий вид [1.47]:

$$\tau(F) = \tau_{min} + \frac{100}{F} \cdot (\tau_{100} - \tau_{min}),$$

где: $\tau(F)$ – значение коэффициента «спада», учитывающего влияние частоты на крутизну спада кривой постмаскировки; τ_{min} – минимальное значение этого коэффициента, в соответствии с [1.47], причем τ_{min} = 8 мс; τ_{100} – значение коэффициента $\tau(F)$ для сигнала частотой 100 Гц; при этом τ_{100} = 30 мс. На рис. 1.54 представлено графическое представление данной функции. По оси ординат отложены значения коэффициента $\tau(F)$, мс; по оси

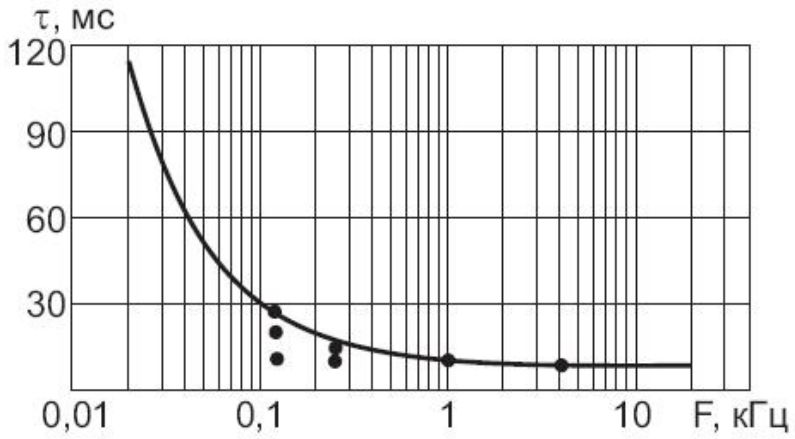


Рис. 1.54. Изменение коэффициента «спада» кривой постмаскировки в зависимости от частоты

абсцисс – значения частоты F, кГц. Кружочками здесь показаны значения коэффициентов «спада» кривых постмаскировки, рассчитанные на базе экспериментальных данных для различных экспертов [1.30]. Видно (рис. 1.54), что с ростом частоты скорость спада кривых постмаскировки уменьшается.

Часто для реализации кривых постмаскировки применяют цифровые фильтры, с помощью которых кривые постмаскировки также реализуются достаточно просто.

1.8. Локализация действительных источников звука

Следует различать действительные и кажущиеся источники звука. Для каждой из этих ситуаций разработаны модели локализации, имеющие как общие, так и индивидуальные особенности.

Известно, что решающую роль в оценке направления на источник звука в реверберирующем звуковом поле играет эффект предшествования или эффект Хааса (*H.Haas*, 1949). Суть его состоит в отделении слуховой системой сигналов прямого звука от их реверберационных продолжений. При этом суждение о направлении на источник звука формируют сигналы прямых звуков, в то время как часть следующих за ними запаздывающих повторений на интервале времени от 1,5 до 30… 50 мс подавляется слуховой системой. При больших временных сдвигах такого подавления не происходит. В этом случае отраженные сигналы рассматриваются как помеха, но и одновременно с этим, по мнению многих авторов, эта часть реверберационного процесса помещения играет важную роль при оценке расстояния до источника звука.

Оценка азимута источника звука. Предположим, что под некоторым углом к медианной плоскости *I-I* головы слушателя находится источник звука Гр (рис. 1.55,*а*). Вследствие дифракции звуковой волны вокруг головы слушателя и частотно-зависимого затухания последней с расстоянием *l* сигналы, приходящие к левому 1 и правому 2 ушам слушателя,

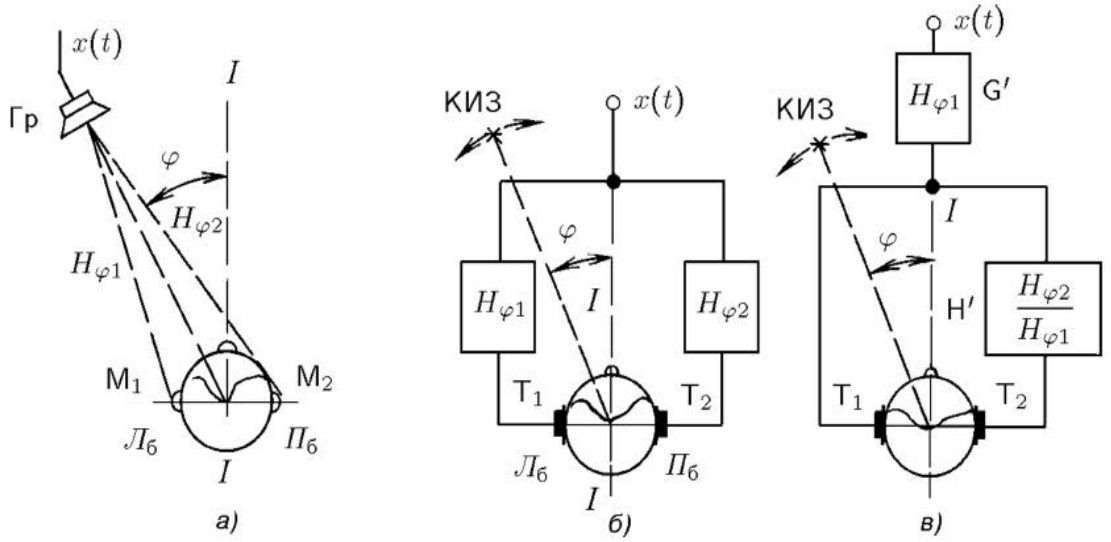


Рис. 1.55. Кодирование места действительного источника звука в пространстве: *а* – к возникновению различий бинауральной пары сигналов; *б* – эквивалентная схема, поясняющая механизм пространственного кодирования сигнала источника; *в* – простейший бинауральный регулятор направления

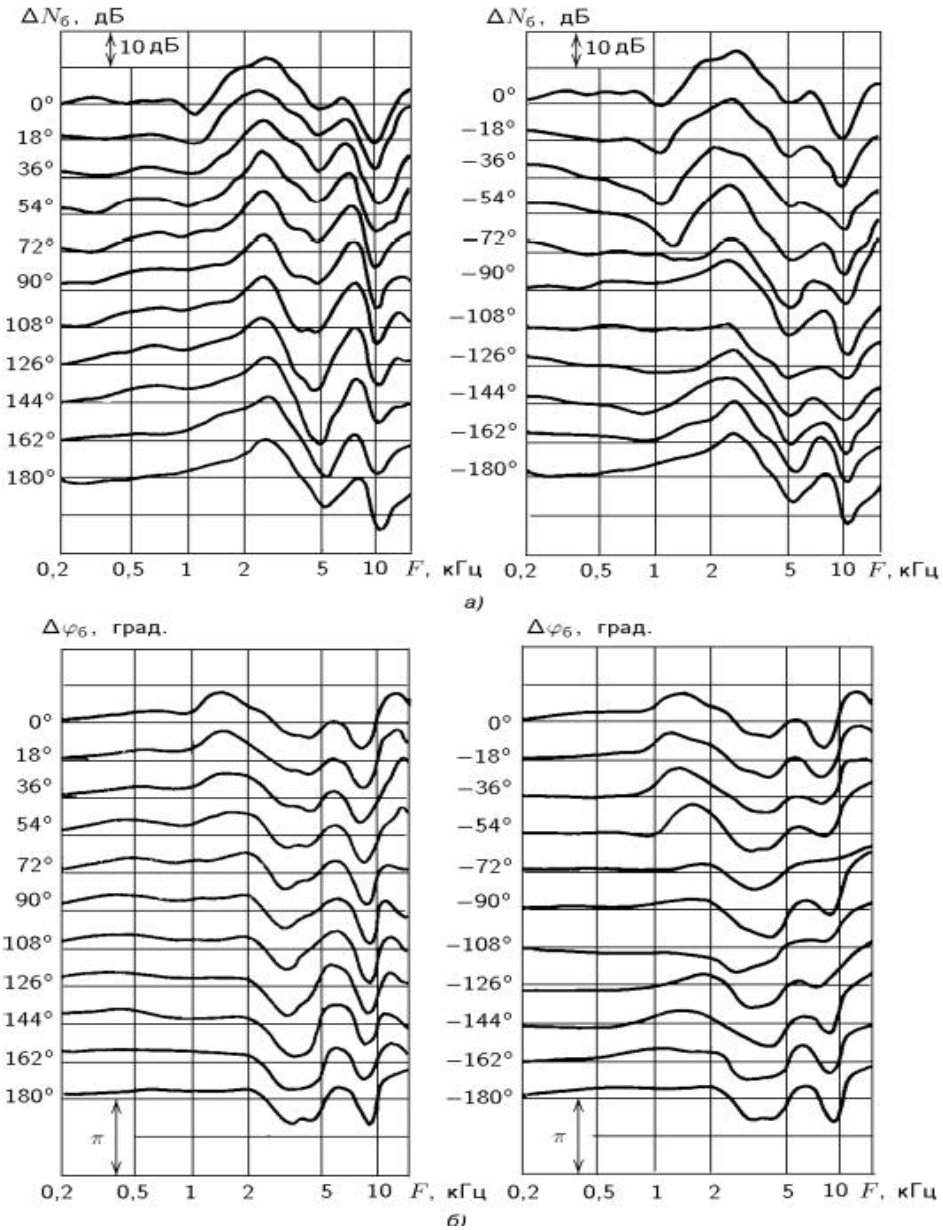


Рис. 1.56.Изменение разности амплитуд (*а*) и фаз (*б*) бинауральной пары сигналов от частоты для разных направлений на действительный источник звука; параметром каждой кривой является величина угла между медианной плоскостью *I-I* головы слушателя и направлением на источник звука

оказываются неодинаковыми. Они отличаются по уровню $\Delta N_б(\varphi,F)$, по времени $\Delta\tau_б(\varphi,F)$ и являются функцией азимута φ и частоты F. В качестве иллюстрации сказанного на рис. 1.56 показаны зависимости, характеризующие разность уровней $\Delta N_б$, в дБ, и разность фаз $\Delta\varphi_б$, в град, бинауральной пары сигналов $\Delta N_б$ и $\Delta\tau_б$ от частоты F, в кГц. Параметром представленных кривых является величина угла φ источника звука Гр относительно медианной плоскости *I-I* (см. рис. 1.55,*а*). Каждому значению φ соответствует своя индивидуальная пара кривых. Значения $\Delta N_б$ и $\Delta\tau_б$ бинау-

ральной пары сигналов, соответствующих данному источнику звука, и являются носителями информации о направлении. Пара сигналов, воздействующая на уши слушателя, может рассматриваться как результат кодирования места источника звука в пространстве. Другими словами, голова и ушные раковины слушателя играют роль пространственных фильтров, а бинауральная пара сигналов на их выходе несет информацию о месте источника звука в пространстве.

Все же пары значений $\Delta N_б$ и $\Delta\tau_б$ не позволяют однозначно оценить азимутальный угол φ источника звука относительно медианной плоскости *I—I* головы слушателя. Действительно (рис. 1.57,*а*) для каждой гиперболы, построенной так, что ее фокусами являются входы 1или 2.

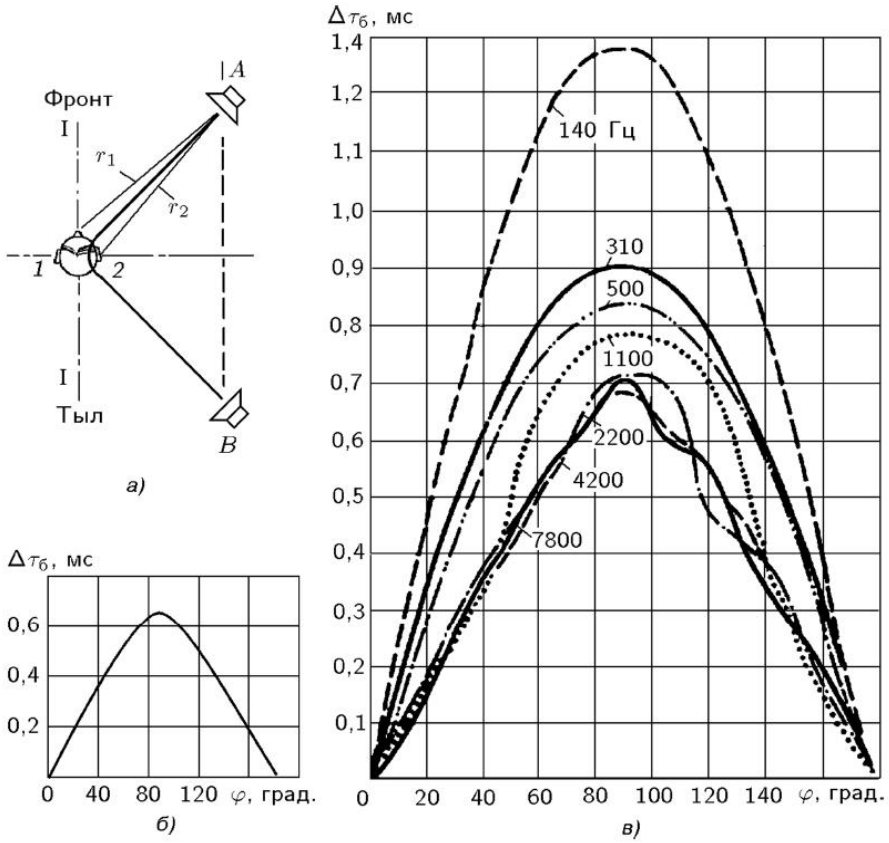


Рис. 1.57.Изменение временного сдвига пары сигналов $Л_б$ и $П_б$ при бинауральном слушании: а– к неоднозначности оценки азимута источников звука *A* и *B*; *б* – влияние частоты испытательного сигнала (цифры у кривых – значение средней частоты 1/3 октавной полосы белого шума) и *в* – среднестатистическая зависимость $\Delta\tau_б$ от φ

и 2 органа слуха, существует всегда множество пар точек (*A*, *B)*, расположенных зеркально относительно линии 1—2, для которых обеспечиваются приблизительно одинаковые значения $\Delta N_б$ и $\Delta\tau_б$ бинауральных сигналов. Например, для источников звука, расположенных в ме-

дианной плоскости *I—I* на одинаковом расстоянии от центра головы слушателя, значения $\Delta N_б$ и $\Delta\tau_б$ примерно одинаковы для фронтального и тылового направлений. Несмотря на это, локализация звуковых образов оказывается безошибочной за счет дополнительного спектрального анализа бинауральной пары сигналов.

Орган слуха человека имеет два механизма для оценки местоположения источника звука в пространстве. Один из них (*фронт-тыл*) определяет, находится ли источник звука спереди или сзади слушателя (относительно линии 1—2), а другой — направление φ на источник звука относительно медианной плоскости (*I—I*, рис. 1.57,*а*). Известно, что значение временной разности $\Delta\tau_б$ бинауральной пары сигналов определяется формулой

$$\Delta\tau_б = \frac{d_{эк}}{c}\sin\varphi = \frac{d\nu(F)}{c}\sin\varphi \ ,$$

где d — база приемников слуховой системы, равная 21 см; c = 340 м/с—скорость распространения фронта звуковой волны; φ — азимут источника звука относительно медианной плоскости; $\nu(F)$ — коэффициент, учитывающий частотно-зависимое влияние ушной раковины и действие последней как линии задержки, время запаздывания фронта звуковой волны в которой зависит от азимута φ источника звука; $d_{эк}$ — эквивалентный размер базы приемников слуховой системы —расстояние между фазовыми центрами раскрыва ушных раковин.

Профессором Я.В.Альтманом [1.52] высказано предположение, что зависимость от азимута $\Delta\tau_б$ является функцией, близкой к линейной:

$$\Delta\tau_б = m_1\varphi, \quad если \quad \begin{matrix} 0^0 < \varphi < 80^0 \\ 100^0 < \varphi < 180^0 \end{matrix} ,$$

где m_1 – постоянный коэффициент. Подтверждением этому являются экспериментальные зависимости $\Delta\tau_б = f_1(\varphi)$, взятые из его же работы и представленные соответственно на рис. 1.57,*б* и *в*, и зависимость смещения КИЗ от величины интераауральной временной разности $\Delta\tau_б$ сигналов, подводимых к левому и правому ушам слушателя с помощью головных телефонов (рис. 1.58,а). Величина φ углового смещения КИЗ пропорциональна значению $\Delta\tau_б$ в диапазоне 0...0,63 мс. При $\Delta\tau_б > 0{,}63$ мс источник звука полностью латерализован, т. е. находится вблизи уха, на которое подается опережающий сигнал.

Значения ΔN и $\Delta\tau$ при локализации взаимозаменяемы, поэтому, если зависимость $\Delta\tau_б = f_1(\varphi)$ является линейной, то и зависимость $\Delta N_б = f_2(\varphi)$

должна быть также линейной функцией от азимута φ источника звука $\Delta N_б = m_2\varphi$. Правильность этого заключения может быть дополнительно подтверждена следующими соображениями. Разность амплитуд

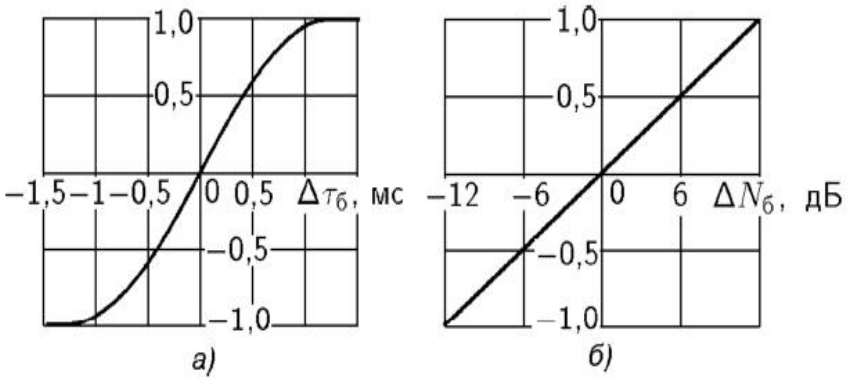


Рис. 1.58. Смещение кажущегося источника звука под действием интерауральных временной (*а*) и интенсивностной (*б*) разностей бинауральной пары сигналов

ΔA бинауральной пары сигналов с учетом характеристик направленности левого и правого ушей слушателя можно найти из выражения

$$\Delta A = m_2^{'} 2\mu \cos(\Delta\varphi - \pi/2)\sin\varphi,$$

где $\Delta\varphi$ – азимут максимума характеристики направленности ушной раковины, отсчитываемой от медианной плоскости, аппроксимируется формулой

$$\Delta\varphi = \frac{1}{12}\pi[4\exp(-0,5\cdot 10^{-6}F^2\delta^2)+5],$$

где $m_2^{'}$ – постоянный коэффициент; $\mu = 0,2\ln(F/F_0)$; F_0=50 Гц – коэффициент, учитывающий изменение характеристики направленности (ХН) ушной раковины с частотой; δ = 1 с.

С учетом изложенного отношение разности амплитуд бинауральных сигналов к их сумме определится как

$$\delta A = \frac{A_1 - A_2}{A_1 + A_2} = \frac{\mu\sin\varphi\cos(\Delta\varphi - \pi/2)}{1+\mu\cos\varphi\sin(\Delta\varphi - \pi/2)}.$$

Данное выражение представляет собой линейную зависимость для значений углов φ, лежащих в пределах 15…90^0, при условии $\Delta\varphi \neq 90^0$, поэтому

$$\delta A = m_2 \varphi \ .$$

Если при оценке азимута φ слух учитывает отношение амплитуд A_1 и A_2 бинауральных сигналов, то

$$\frac{A_1 - A_2}{A_1 + A_2} = m\varphi \ , \qquad a \qquad \frac{A_1}{A_2} = \frac{1 + m\varphi}{1 - m\varphi} \ ,$$

поэтому отношение A_1/A_2 также линейно зависит от φ. Отличие состоит лишь в изменении угла наклона этих зависимостей. Здесь левая часть равенств выражена в децибелах.

Этот вывод подтверждается и экспериментальными данными. На рис. 1.58,б приведена зависимость смещения КИЗ от величины бинауральной интенсивностной разности $\Delta N_б = 20\lg(A_1/A_2)$ сигналов, воспроизводимых головными телефонами. На рис. 1.59 представлены результаты измерений величин $\Delta N_б$.

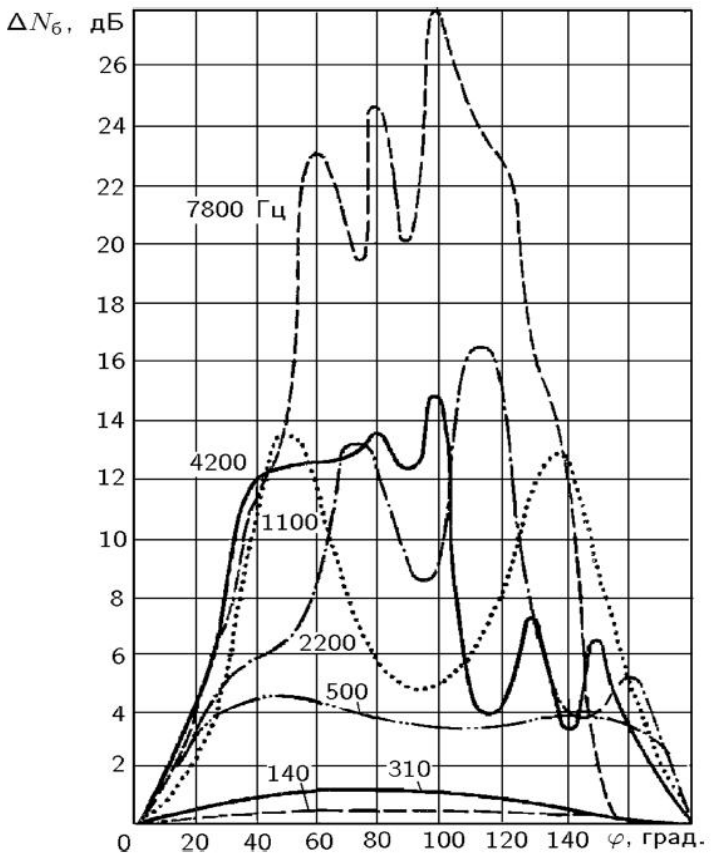


Рис. 1.59. Изменение интенсивностной разности $\Delta N_б$ бинауральной пары сигналов $Л_б$ и $П_б$ от азимута источника звука (цифры у кривых – значение средней частоты 1/3-октавной полосы белого шума)

Взаимозаменяемость значений $\Delta\tau_б$ и $\Delta N_б$ при локализации звуковых образов позволяет ввести понятие коэффициента эквивалентности $K_б$, дБ/мс для пары бинауральных сигналов и определить его как отношение величин $\Delta N_б$ и $\Delta\tau_б$, вызывающих одинаковое смещение источника или взаимно компенсирующих друг друга:

$$K_б=(\Delta N_б / \Delta\tau_б).$$

Поочередное предъявление стимулов $\Delta N_б$ и $\Delta\tau_б$ дает значение $K_б \approx 13$ дБ/мс. Эта величина близка к значению, найденному для обычной стереофонии (~10 дБ/мс).

Если допустить, что орган слуха при оценке азимута φ источника звука обменивает $\Delta\tau_б$ на эквивалентное значение интенсивностной разности $(\Delta N_{эк})_б = K_б\Delta\tau$ вследствие явления торможения в слуховой системе, то справедлива запись

$$(\Delta N_{\Sigma эк})_б = \Delta N_б + K_б\Delta\tau_б.$$

Здесь $(\Delta N_{\Sigma эк})_б$ - суммарное значение эквивалентной интенсивностной разности, вызывающей то же самое смещение источника звука, что и одновременно действующие величины $\Delta N_б$ и $\Delta\tau_б$.

Если учесть, что значения $\Delta\tau_б$ и $\Delta N_б$ являются линейными функциями азимута φ источника звука ($\Delta N_б=m_2\varphi$ и $\Delta\tau_б=m_1\varphi$), а величина коэффициента эквивалентности не должна зависеть от φ ($K_б$ = const), то, очевидно, что величина суммарной эквивалентной интенсивностной разности также является линейной функцией азимута φ, т. е.

$$(\Delta N_{\Sigma эк})_б = m\varphi,$$

где m — постоянный коэффициент.

Итак, каждому значению азимута φ источника звука соответствует пара значений $\Delta\tau_б$ и $\Delta N_б$ или одно значение $(\Delta N_{\Sigma эк})_б$. Орган слуха, по-видимому, использует обе эти возможности для оценки направления.

Заметим, что $\Delta\tau_б$ действует всегда в согласии с $\Delta N_б$ при локализации действительного источника звука. Предполагается, что значение параметра $\Delta\tau_б$ вычисляется слуховой системой по максимуму взаимной корреляционной функции бинауральной пары сигналов.

Величины $\Delta N_б$ и $\Delta\tau_б$ являются не только линейными функциями азимутального угла, но и зависят от частоты (рис. 1.57,*в* и 1.59). Они изменяются при переходе от одной частотной группы слуха к другой, оставаясь, по-видимому, примерно постоянными внутри нее. Однако, величина $(\Delta N_{\Sigma эк})_б$

при переходе от одной частотной группы слуха к другой при $(\Delta N_{\Sigma эк})_б$ = const по-видимому изменяться не должна, так как объем слуховой памяти ограничен.

На низких частотах (ниже 500 Гц) $\Delta N_б << K_б \Delta\tau_б$ и оценка азимута практически определяется только значением $\Delta\tau_б$; в диапазоне средних частот (500...5000 Гц) оба фактора $\Delta\tau_б$ и $\Delta N_б$ приблизительно в равной степени способствуют созданию ощущения направления. На высоких частотах (выше 5000 Гц) $\Delta N_б >> K_б \Delta\tau_б$, т.е. оценка азимута практически определяется величиной $\Delta N_б$. На частотах ниже 150 Гц локализация источника звука невозможна. Сохранение параметра $(\Delta N_{\Sigma эк})_б$ неизменным при переходе от одной частотной группы слуха к другой для постоянного значения азимута φ возможно, если $K_б$ будет являться функцией частоты. При этом изменение коэффициента эквивалентности должно компенсировать частотную зависимость $\Delta N_б$ и $\Delta\tau_б$.

Временной $\Delta\tau_б$ и интенсивностный $\Delta N_б$ факторы действуют на орган слуха независимо. Поэтому значение $K_б$ может быть легко найдено методом компенсации. Предварительно введением в пару бинауральных сигналов, например, значения $\Delta N_б$ латерализуют слуховой КИЗ, а затем с помощью $\Delta\tau_б$ возвращают его на прежнее место (медианная плоскость). В этом случае действие одного фактора компенсируется влиянием другого. Путем проведения тщательных экспериментальных исследований (на полосах белого шума) было установлено, что при компенсации величина коэффициента эквивалентности $K_б$ изменяется в пределах от 5 до 30…50 дБ/мс и зависит от средней частоты испытательного сигнала, что подтверждает высказанное выше соображение: $(\Delta N_{\Sigma эк})_б$ = const при переходе от одной критической полосы слуха к другой, если азимутальное положение источника звука остается неизменным.

Все изложенное позволяет описать работу механизма локализации слуха при оценке азимута источника звука следующим образом. Голова и ушные раковины слушателя играют роль пространственного фильтра, осуществляющего пространственное кодирование сигналов, поступающих от источника звука к левому и правому ушам слушателя. Полученная в результате пространственного кодирования пара бинауральных сигналов содержит всю необходимую информацию для оценки местоположения источника звука в пространстве: угловое смещение от медианной плоскости, расположение спереди или сзади слушателя, возвышение над горизонтальной плоскостью, удаление.

Суждение о величине углового смещения φ источника звука от медианной плоскости связано с оценкой слуховой системой временных $\Delta\tau_б$ и интенсивностных $\Delta N_б$ различий пары бинауральных сигналов, а также и величины $(\Delta N_{\Sigma эк})_б$. Полученные в результате пространственного кодирования величины $\Delta N_б$ и $\Delta\tau_б$, а также и вычисленное значение $(\Delta N_{\Sigma эк})_б$ сравниваются в каждой частотной группе слуха с заученными (приоб-

ретенными в результате опыта) эталонными образцами, хранящимися в слуховой памяти. Идентификация (частичная или полная) «измеренной» пары значений $\Delta N_б$, $\Delta\tau_б$ и вычисленной величины $(\Delta N_{Σэк})_б$ с одним из хранящихся в памяти образцов позволяет слушателю оценить величину углового смещения источника звука в пространстве относительно медианной плоскости. Неоднозначность оценки «фронт—тыл» устраняется путем частотного анализа бинауральных стимулов.

Работа механизма «фронт-тыл». Ключевым моментом для понимания работы механизма «*фронт— тыл*» является зависимость, показанная на рис. 1.60,*а*. Она представляет собой изменение разности уровней звукового давления:

$$\Delta N_{ФТ} = N_Ф - N_Т,$$

где $N_Ф$ - уровень звукового давления, создаваемый у барабанной перепонки фронтальным громкоговорителем; $N_Т$ - то же самое, но для тылового громкоговорителя. В обоих случаях источник звука Гр находится

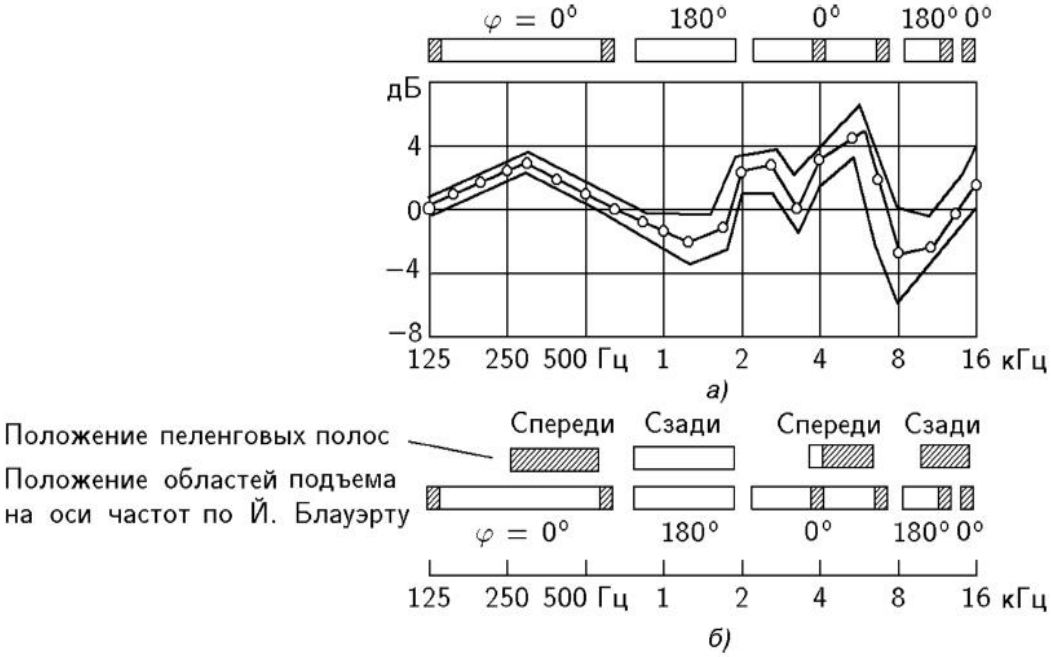


Рис. 1.60. К пояснению особенностей работы механизма «*фронт – тыл*»: *а* – изменение разности звукового давления фронтального и тылового громкоговорителей от частоты; *б* – расположение полос направления на оси слышимых частот

в медианной плоскости *I-I* на одинаковом расстоянии от центра головы слушателя. Из рис. 1.60,*а* следует, что в отдельных частотных областях фронтальный источник звука создает большее звуковое давление: $N_Ф > N_Т$, в других частотных полосах наблюдается обратное явление: $N_Т > N_Ф$. Более глубокое изучение фильтрующего действия головы и ушных рако-

вин слушателя позволило ввести понятие так называемых пеленговых полос или полос направления [1.64;1.65; 1.66]. Их расположение на оси частот показано на рис. 1.60,*б*. Видно, что пеленговые полосы, соответствующие расположению источника звука спереди ($\varphi = 0°$), связаны с областями частот, где $N_Ф > N_Т$. Пеленговые полосы, соответствующие тыловому направлению ($\varphi = 180°$), связаны с частотными областями, где $N_Т > N_Ф$. Для уверенной фиксации слушателем фронтального или тылового направления достаточно иметь разбаланс громкоговорителей по уровню $|N_Ф - N_Т| > 1{,}5...2$ дБ. Предполагается, что ощущение направления «фронт—тыл» формируется преимущественно теми полосами направления, в которых сосредоточена большая часть энергии сигнала. Заметим, что условия работы механизма «фронт—тыл» должны ухудшаться для источников звука, находящихся вне медианной плоскости $\varphi \neq 0°$ и $\varphi \neq 180°$. Этот механизм не работает при $\varphi = 90°$ или $\varphi = 270°$.

Вполне возможно, что при оценке направления «фронт—тыл» слуховой системой также учитывается тот факт, что ушные раковины играют роль линии задержки, временной сдвиг которой является функцией азимута φ источника звука. Результат пространственного кодирования места источника звука сравнивается с эталонными для каждого направления образцами. Суждение о направлении является следствием идентификации результатов анализа бинауральной пары сигналов с одним из эталонных образцов.

Оценка угла возвышения источника звука. До сих пор мы говорили исключительно о бинауральной оценке азимута источника звука. В отличие от изложенного признаки, лежащие в основе оценки угла возвышения источника звука, часто считают моноуральными. Ушная раковина действует подобно акустической антенне. Ее резонансные полости усиливают некоторые частоты, а ее геометрия приводит к интерференции волн, которая уменьшает другие частоты. Кроме того, частотная характеристика уха зависит от направления прихода звуковой волны (рис. 1.61,*а*). В каждом случае имеются два пути распространения звука от источника до канала уха: прямой путь распространения звуковой волны и более длинный, на котором волна претерпевает отражение от ушной раковины. На умеренно низких частотах, ушная раковина по существу собирает дополнительную звуковую энергию, и сигналы этих двух путей приходят в фазе. Однако на высоких частотах, задержанный сигнал не совпадает по фазе с прямым сигналом, и происходит их взаимное ослабление. Самое большое ослабление происходит, когда разность в длине пути d равна половине длины волны, то есть, когда $F = c/2d$. В показанном примере, это создает впадину на АЧХ – минимум вокруг частоты 10 кГц. Для типичных значений d, частота минимума обычно лежит в диапазоне от 6 до 16 кГц. Так как ушная раковина более эффективный рефлектор для звуков, приходящих спереди, чем сверху, результирующий минимум намного более заметен для источников

находящихся спереди, чем сверху. Кроме того, разность длин путей меняется с углом возвышения, поэтому частота минимума также движется с возвышением. Хотя все еще имеются споры относительно того, какие особенности являются наиболее важными для оценки угла возвышения, все же установлено, спектральное окрашивание звука, определяемой ушной раковиной, обеспечивает первичные признаки возвышения. Его пример показан на рис. 1.61,*б*. Источник звука был расположен в двух метрах слева от слушателя и перемещался от уровня уха (0 градусов) до возвышения в 30 градусов над уровнем уха (сплошная линия – 0 градусов; длинная штриховая - 10 градусов; короткая штриховая – 20 градусов; пунктирная – 30 градусов) [1.67].

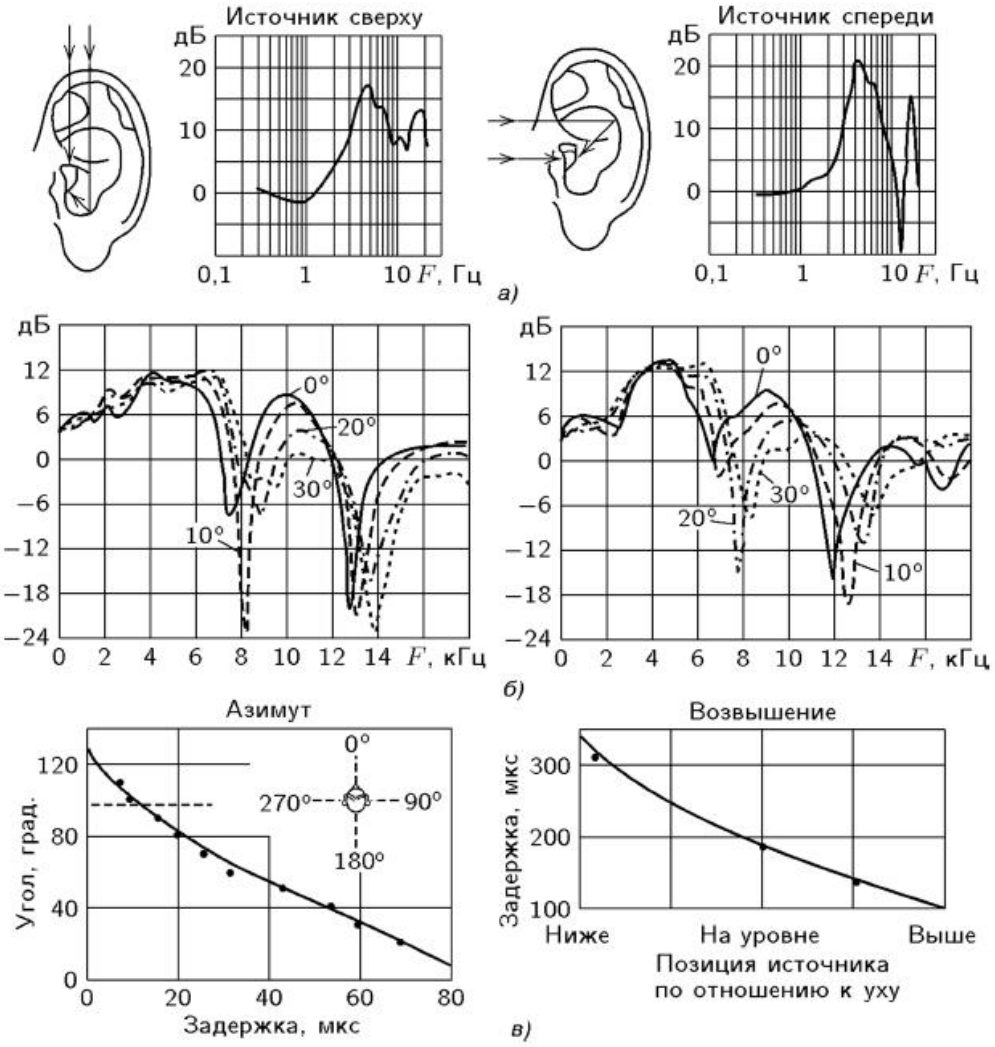


Рис. 1.61.Траектории попадания в ушной канал звуковой волны от источника звука (*а*), спектральное окрашивание (*б*) и задержки, возникающие при отражении звуковой волны от краев ушной раковины (*в*)

В ряде работ утверждается также, что весьма важную роль играют здесь и задержки в приходе звуковых волн, отраженных от ушной раковины (рис. 1.61,в). Левый рисунок здесь показывает задержку (мкс), возникаю-

щую при отражении от краев внутреннего уха, которые определяют различия фронт – тыл в горизонтальной плоскости. Правый рисунок показывает задержку при отражении от внешнего края ушной раковины, которые важны при определении возвышения источника в вертикальной плоскости [1.68].

Глубинная локализация. Наряду с азимутом слушатель также достаточно уверенно оценивает и расстояние l до источника звука. Перечислим признаки бинауральной пары сигналов, оказывающих влияние на оценку параметра l.

1. При средних значениях l от 3 до 15…20 м приближение и удаление источника звука сопровождаются заметным изменением его интенсивности. В свободном звуковом поле увеличение расстояния до источника звука в 2 раза сопровождается уменьшением уровня звукового давления на 6 дБ. Экспериментальные данные подтверждают связь оценки расстояния l с уровнем интенсивности источника звука (N, в дБ). В качестве примера на рис. 1.62 приведена соответствующая зависимость, заимствованная из работ проф. А.Я.Альтмана. Однако, чтобы использовать громкость для определения расстояния до источника звука, нам

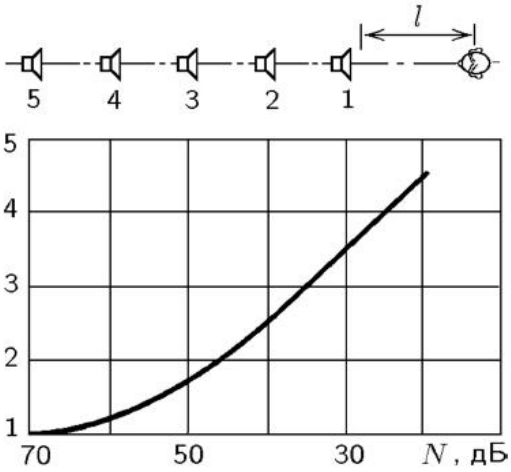


Рис. 1.62 Кажущаяся удаленность источника звука в зависимости от его интенсивности

необходимо также знать кое-что относительно характеристик источника звука. В случае человеческой речи, каждый из нас знает из личного опыта различное качество звука, соответствующее шепоту, нормальному разговору и крику, независимо от уровня звука. Комбинация громкости и знания источника дает нам полезную информацию для оценки расстояния до источника звука.

2. При малых расстояниях l до источника звука (l<2 м) наблюдаются изменения спектра сигналов вследствие искажения фронта звуковой волны головой и ушными раковинами. При l, превышающих 10…15 м, начинает сказываться частотно-зависимое затухание звуковой волны в

воздухе с расстоянием. Оба вида изменений формы спектра влияют на оценку расстояния до источника звука.

Кроме того, возрастание амплитуды низкочастотных составляющих в спектре сигнала связано с ощущением приближения источника звука; искусственное уменьшение амплитуды высокочастотных составляющих в спектре воспринимается как удаление источника звука.

3. В отличие от азимутальной глубинная локализация возможна и при моноуральном слушании, но бинауральное восприятие существенно повышает точность оценки параметра l. Орган слуха, оценивая величины $\Delta\tau_б$ и $I_{ср}/\Delta I_б$, определяет расстояние до источника звука

$$l = 2c\Delta\tau_б(I_{ср} / \Delta I_б),$$

где $I_{ср}$ – среднее значение интенсивностей сигналов, воздействующих на уши слушателя; $\Delta I_б$ – бинауральная разность интенсивностей; c – скорость звука.

Теоретический анализ этого выражения показывает, что при $l > 10$ м необходимо предъявлять очень жесткие требования по разрешающей способности временных интервалов и приращений интенсивности, значительно превышающие возможности человека. Однако при значениях l порядка единиц метров необходимая разрешающая способность находится в пределах, доступных человеческому уху. Этот способ оценки параметра l может играть существенную роль в условиях открытого пространства или заглушенной камеры.

4. В помещениях, где наряду с прямым звуком на слушателя воздействует значительное число отраженных волн, важным фактором, стимулирующим глубинную локализацию, является реверберация, точнее, величина акустического отношения. Благодаря эффекту предшествования слуховой анализатор способен оценить энергию прямых звуков и отзвуков, составляющих реверберационный процесс. Используя известное выражение для акустического отношения, можно записать

$$l = \sqrt{\frac{\varepsilon_1}{\varepsilon_2}\frac{\alpha Q_\Sigma}{50(1-\alpha)}},$$

где $\varepsilon_1/\varepsilon_2$ – отношение плотностей энергий отраженных и прямого звука, известное под названием акустического отношения, α – средний коэффициент звукопоглощения, Q_Σ – площадь поверхностей помещения.

Большинство исследователей считают этот фактор важнейшим при оценке расстояния l. Все же необходимо признать, что стройной модели, объясняющей с достаточной полнотой механизм оценки расстояния l, пока нет. Накопленные здесь сведения следует считать как весьма скромные.

Слушательский опыт свидетельствует о том, что глубинная локализация в естественных условиях не отличается большой точностью.

1.9. Локализация кажущихся источников звука

Условие образования кажущегося источника звука. Рассмотрим особенности образования и локализации звуковых образов на примере двухканальной стереофонии. Предположим, что слушатель находится на оси симметрии *Y* системы воспроизведения $Гр_1$ и $Гр_2$, а излучаемые громкоговорителями сигналы Л и П не имеют различий по времени ($\Delta\tau = 0$) и уровню ($\Delta N = 0$) и получены от одного и того же источника звука *M* (рис. 1.63).

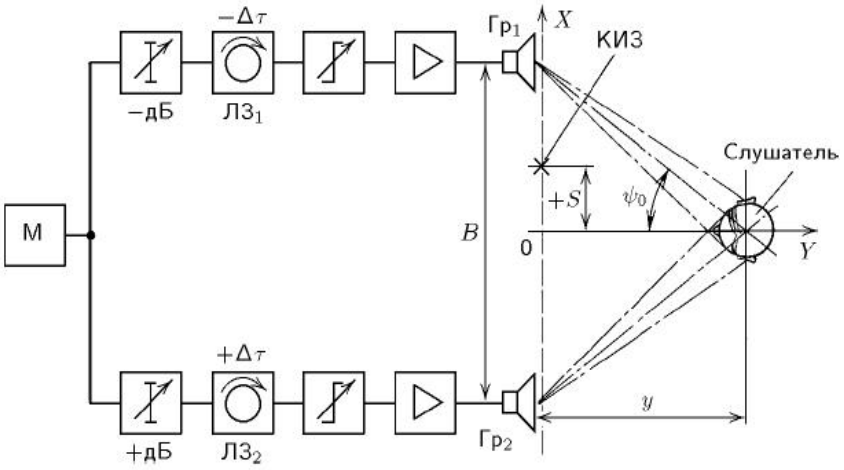


Рис. 1.63. Экспериментальная установка для изучения особенностей локализации КИЗ: $ЛЗ_1$ и $ЛЗ_2$ – линии задержки; М – магнитофон

Громкоговорители $Гр_1$ и $Гр_2$ включены синфазно. В этом случае звучания обоих громкоговорителей сливаются в единый звуковой образ, который кажется слушателю расположенным посередине линии базы громкоговорителей в точке 0. Этот звуковой образ является кажущимся, его появление возможно, если сигналы, излучаемые громкоговорителями, статистически связаны (коррелированны). По мере снижения коэффициента корреляции между канальными сигналами КИЗ локализуется все менее четко, его протяженность увеличивается, и при уменьшении коэффициента корреляции $R(\Delta\tau)$ сигналов Л и П до значения 0,05...0,15 наступает разрыв КИЗ на два действительных источника звука. Последние воспринимаются раздельно и локализуются соответственно в позициях левого ($Гр_1$) и правого ($Гр_2$) громкоговорителей.

Феномен образования КИЗ, возможность его локализации в разных точках пространства – наиболее яркая особенность стереовоспроизведения, определяющая такой его признак качества, как пространственное впечатление. Локализация КИЗ включает оценку азимута и расстояния до источ-

ника звука. Глубинная локализация КИЗ изучена недостаточно: не в полной мере выявлены стимулирующие ее факторы, не разработаны модели этого механизма слуха, адекватные восприятию.

Положение КИЗ на линии базы громкоговорителей (оценка его азимута) зависит только от временных и интенсивностных различий между сигналами, достигающими ушей слушателя. Эти различия могут быть обусловлены либо свойствами сигналов Л и П стереопары (ΔN и $\Delta\tau$), либо местом расположения слушателя относительно громкоговорителей ($\Delta N_{x,y}$, $\Delta\tau_{x,y}$, где x и y – координаты слушателя). При дальнейшем рассмотрении смещение КИЗ (S, см. рис. 1.63) вправо от центра базы будем считать положительным (+S), а влево – отрицательным (- S).

Интенсивностная стереофония ($\Delta\tau$=0, $\Delta N \neq 0$), симметричное расположение слушателя относительно громкоговорителей (x = 0, $y \neq 0$). Зависимости, характеризующие смещение КИЗ под действием разности уровней ΔN сигналов Л и П, для разных условий проведения эксперимента представлены на рис. 1.64. Здесь по оси ординат отложено относительное смещение $S/(B/2)$ кажущегося источника звука в долях полубазы $B/2$, а по оси абсцисс – значения ΔN в децибелах, определяемые как 20 $\lg(p_{3B2}/p_{3B1})$, где p_{3B1} и p_{3B2} – звуковые давления, развиваемые соответственно правым и левым громкоговорителями СВ.

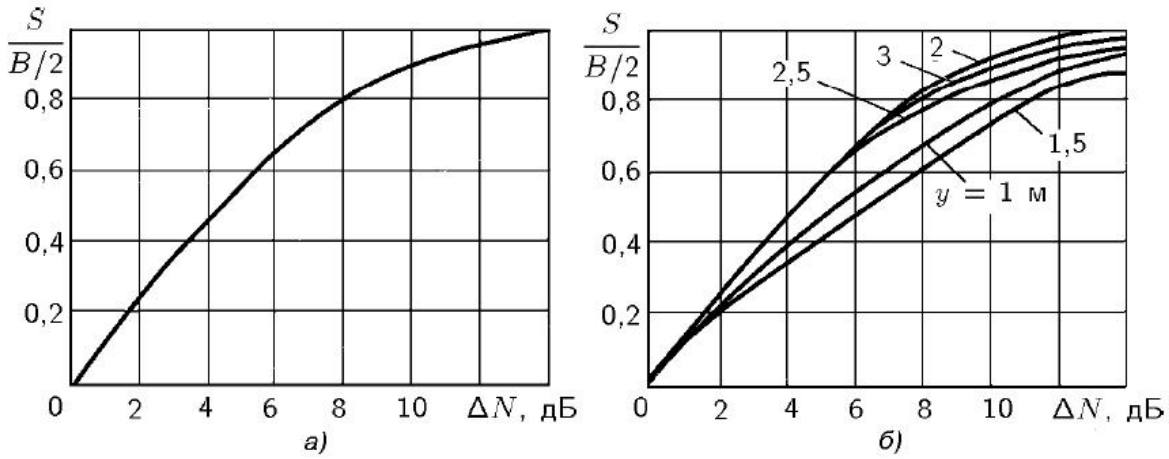


Рис. 1.64. Зависимость относительного смещения КИЗ от разности уровней при $\Delta\tau = 0$ и $x = 0$: 1 – y=1 м; 2 – y = 1,5 м; 3 – y = 2,5 м; 4 – y = 3 м; 5 – y = 2 м

Характер зависимостей $S/(B/2) = f_1(\Delta N)$ на рис. 1.64 для всех типов вещательных сигналов (речевых и музыкальных) одинаков. Введение ΔN сопровождается перемещением КИЗ от своего первоначального положения (S = 0 при ΔN = 0) в сторону громкоговорителя, излучающего сигнал с большим уровнем. При ΔN = 12...16 дБ КИЗ локализуется практически в позиции громкоговорителя и дальнейшее увеличение ΔN не вызывает его дальнейшего перемещения.

Величина относительного смещения $S/(B/2)$ КИЗ при ΔN = const не зависит от расстояния y между слушателем и линией базы $Гр_1$ и $Гр_2$ для малых баз (B = 0,8...1,8 м, рис. 1.64,*а*). Поэтому здесь представлены результаты, усредненные по y.

Для относительно больших баз ($B \geq 2{,}8$ м) это утверждение справедливо только при $y \geq B$ (рис. 1.64,*б*). При приближении слушателя к системе воспроизведения на расстояние y < B наблюдается (при ΔN = const) смещение КИЗ тем меньшее, чем ближе расположен слушатель к линии базы громкоговорителей.

Реверберационный процесс помещения прослушивания (из-за эффекта предшествования) практически не влияет на азимутальное положение КИЗ, но приводит к увеличению протяженности последнего (вследствие снижения корреляции между воспринимаемыми сигналами), что снижает точность локализации КИЗ.

Временная стереофония (ΔN=0, $\Delta\tau \neq 0$), симметричное положение слушателя относительно громкоговорителей системы воспроизведения ($x = 0$, $y \neq 0$). При введении временного сдвига $\Delta\tau$ КИЗ смещается в сторону громкоговорителя, излучающего опережающий сигнал. Перемещение КИЗ с увеличением $\Delta\tau$ носит монотонный характер только для сигналов, спектры мощности которых не имеют ярко выраженных неоднородностей распределения энергии по частоте (рис. 1.65,*а*). При изменении $\Delta\tau$ от 0 до

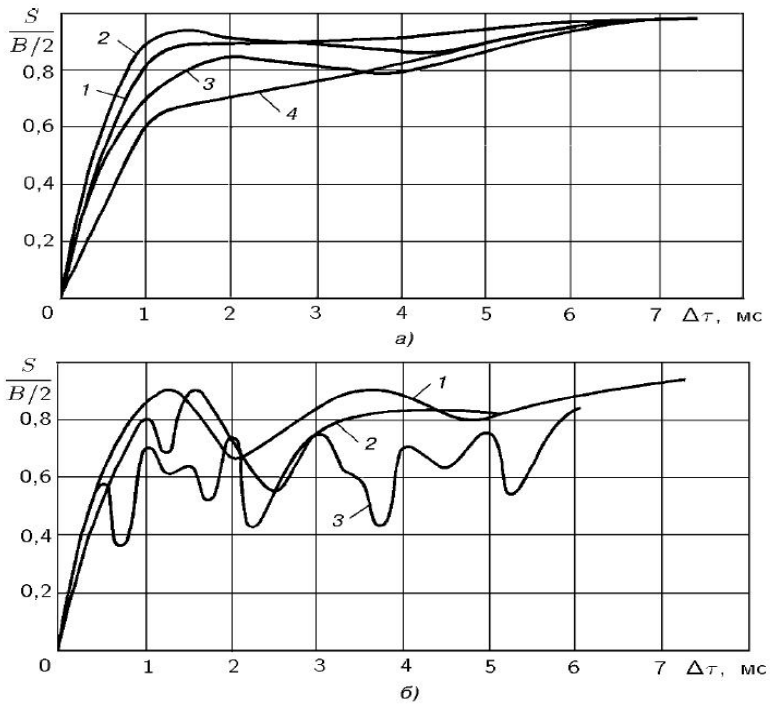


Рис. 1.65. Зависимость относительного смещения КИЗ от временной разности сигналов стереопары с достаточно однородным спектром мощности (*а*, где 1- арфа, 2 – рояль, 3 – труба, 4 – кастаньеты) и неравномерным распределением мощности по частоте (*б*, где 1 – женская речь, 2 – флейта, 3 – скрипка) при B = 1,8 м; ΔN = 0; x = 0; y = 1,5 м

0,8...1,2 мс наблюдается быстрое перемещение КИЗ до (0,7...0,8)B/2. Дальнейшее увеличение $\Delta\tau$ до $\Delta\tau'_{пор}$ = 30...150 мс, соответствующей разрыву КИЗ на два действительных источника звука, сопровождается медленным его перемещением на участке (0,7...0,8)·B/2. Для сигналов, спектры мощности которых имеют ряд энергетических пиков, зависимость $S/(B/2) = f_2(\Delta\tau)$ носит ярко выраженный индивидуальный характер. В этом случае монотонное перемещение КИЗ наблюдается только на начальном участке кривой $S/(B/2) = f_2(\Delta\tau)$, где $\Delta\tau$ < 0,5...1,0 мс (рис. 1.65,*б*). В интервале же временных задержек от 0,5...1,0 до 5...7 мс наблюдается неоднократное возвращение КИЗ к центру базы с ростом $\Delta\tau$. Однако величина этих "колебаний" уменьшается с увеличением $\Delta\tau$ и уже при $\Delta\tau$ = 5...7 мс становится незначительной.

Увеличение временной разности сопровождается уменьшением корреляции между сигналами Л и П, что приводит к ухудшению четкости локализации. С ростом $\Delta\tau$ (при $\Delta\tau$ > (5...7 мс)) в звучании появляется гулкость, растет протяженность КИЗ, которая при $\Delta\tau \approx \Delta\tau'_{пор}$ становится равной B. Распад КИЗ (при ΔN = 0) наступает при коэффициенте корреляции сигналов стереопары $R(\Delta\tau)$ = 0,05...0,1. Значения $\Delta\tau'_{пор}$ для разных типов звуковых сигналов приведены в табл. 1.6.

Таблица 1.6.

Пороговые значения временного сдвига сигналов, вызывающие разрыв кажущегося источника звука

Вид вещательного сигнала	$\Delta\tau'_{пор}$, мс	$\Delta\tau''_{пор}$, мс	Вид вещательного сигнала	$\Delta\tau'_{пор}$, мс	$\Delta\tau''_{пор}$, мс
Кастаньеты	30	4	Речь женская	50	5...8
Ксилофон	50	5	Скрипка	100	9...12
Барабан малый	50	5	Альт	100	9...12
Барабан большой	50	6	Рояль	100	12...16
Коробочка	40	6	Кларнет	100	16...20
Бубен	40	6	Виолончель	100	21... 30
Речь мужская	50	5...7			

Смешанная стереофония ($\Delta N \neq 0$, $\Delta\tau \neq 0$), симметричное положение слушателя относительно громкоговорителей системы воспроизведения. В этом случае оценка азимута КИЗ определяется совместным независимым действием на орган слуха величин ΔN и $\Delta\tau$. Компенсация временного сдвига $\Delta\tau$ разностью уровней ΔN (рис. 1.66, кривая 1) возможна до тех

пор, пока $\Delta\tau < \Delta\tau''_{пор}$. Значение $\Delta\tau''_{пор}$, при котором наступает распад КИЗ, также зависит от структуры сигнала, причем всегда значительно меньше $\Delta\tau'_{пор}$ (см. табл.1.6). При компенсации распад КИЗ наступает при $R(\Delta\tau) \approx$ 0,15...0,2. Совместное действие $\Delta\tau$ и ΔN (кривая 1) сопровождается следующими изменениями в слуховой оценке:

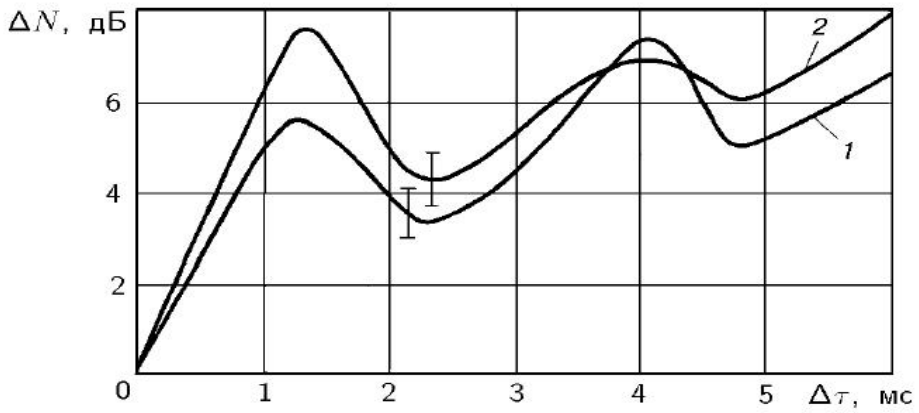


Рис. 1.66. Связь временного ($\Delta\tau$) и интенсивностного (ΔN) факторов (B = 1,8 м; y = 1,8 м и x = 0): 1 – кривая компенсации; 2 – та же самая зависимость, полученная при поочередном действии $\Delta\tau$ и ΔN

1) при $\Delta\tau$ < 3...4 мс опережающий и задержанные сигналы формируют компактный, четкий КИЗ; протяженность КИЗ составляет 10...20 см и не изменяется с введением ΔN. Такое восприятие будем называть устойчивым слиянием опережающего и задержанного сигналов;

2) при 3...4 < $\Delta\tau$ < 7...15 мс локализация КИЗ затруднена, звучание приобретает гулкость и объемность. С ростом $\Delta\tau$ (особенно при попытке компенсации временного фактора разностью уровней) начинает изменяться протяженность КИЗ. Она максимальна, если КИЗ расположен в центре базы громкоговорителей, когда действие $\Delta\tau$ скомпенсировано разностью уровней. Образование КИЗ в данном случае возможно еще при любых значениях $\Delta\tau$. Это области почти устойчивого слияния опережающего и задержанного сигналов;

3) при (7...15) < $\Delta\tau$ < (100...120) мс слияние обоих сигналов и образование КИЗ возможно, если $\Delta N = N_{ЗАП} - N_{ОП} < \Delta N_П$, где $N_{ЗАП}$ и $N_{ОП}$ — уровни запаздывающего и опережающего сигналов; $\Delta N_П$ – пороговое значение ΔN, соответствующее распаду КИЗ. Зависимость $\Delta N_П$ от $\Delta\tau$ представлена на рис. 1.67. Она соответствует речевому КИЗ, получена при B = 2,4 м и $y = B$ (x = 0). При приближении ΔN (компенсирующей действие $\Delta\tau$) к $\Delta N_П$ протяженность КИЗ возрастает и при $\Delta N = \Delta N_П$ становится равной величине базы. Область выше кривой соответствует распаду КИЗ и раздельному восприятию звучаний двух действительных источников звука – громкого-

ворителей. Этот временной интервал – область неустойчивого слияния опережающего и задержанного сигналов;

3) при Δτ > 100...120 мс слияние невозможно, слушатель воспринимает раздельно звучание двух действительных источников звука – громкоговорителей, если превышается порог слышимости.

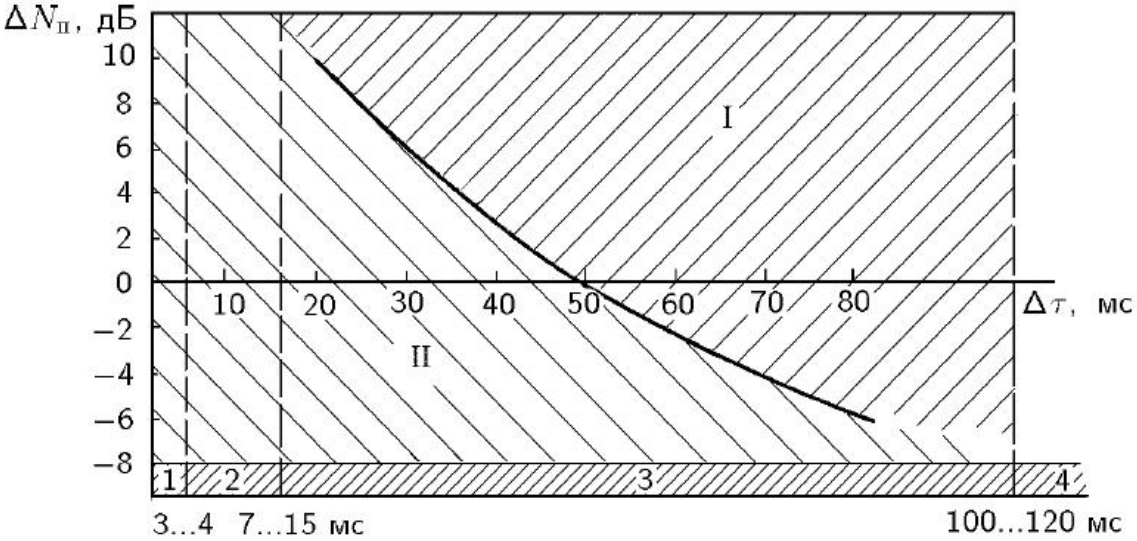


Рис. 1.67. Зависимость порогового значения интенсивностной разности $\Delta N_{П}$, вызывающей разрыв КИЗ, от Δτ: I и II –области соответственно раздельного и слитного восприятия звучаний $Гр_1$ и $Гр_2$

Приведенные выше числовые значения соответствуют речевому сигналу. Для сигналов других типов качественно картина не изменяется, возникают лишь те или иные количественные изменения.

Асимметричное положение слушателя относительно громкоговорителей системы воспроизведения. При боковом смещении x слушателя (рис. 1.68,*а*) появляются дополнительные интенсивностные $\Delta N_{x,y}$ и временные $\Delta\tau_{x,y}$ различия сигналов, поступающих от громкоговорителей в точку прослушивания *A(x, y):*

$$\Delta N_{x,y} = 20\lg\frac{l_1(x,y)D_2(\psi_2)}{l_2(x,y)D_1(\psi_1)}, \Delta\tau_{x,y} = \left[l_1(x,y) - l_2(x,y)\right]/c_{зв},$$

где $l_1(x,y)=\sqrt{(B/2+x)^2+y^2}$, $l_2(x,y)=\sqrt{(B/2-x)^2+y^2}$ – расстояния до $Гр_1$ и $Гр_2$; B – размер базы громкоговорителей; x, y – координаты слушателя; $c_{зв}$ – скорость звука; $D_1(\psi_1)$, $D_2(\psi_2)$ – значения характеристик направленности левого ($Гр_1$) и правого ($Гр_2$) громкоговорителей соответственно для углов ψ_1 и ψ_2.

Оба фактора $\Delta N_{x,y}$ и $\Delta\tau_{x,y}$ действуют в согласии, вызывая смещение КИЗ в сторону ближайшего к слушателю громкоговорителя. Основную роль при

этом играет $\Delta\tau_{x,y}$. Однако нельзя пренебрегать и влиянием $\Delta N_{x,y}$, особенно при малых значениях y и больших значениях x. Влияние бокового смещения слушателя на локализацию КИЗ показано на рис. 1.68,*б*. Перемещение КИЗ начинает ощущаться только при достижении определенного значения ΔN, компенсирующего действие на орган слуха

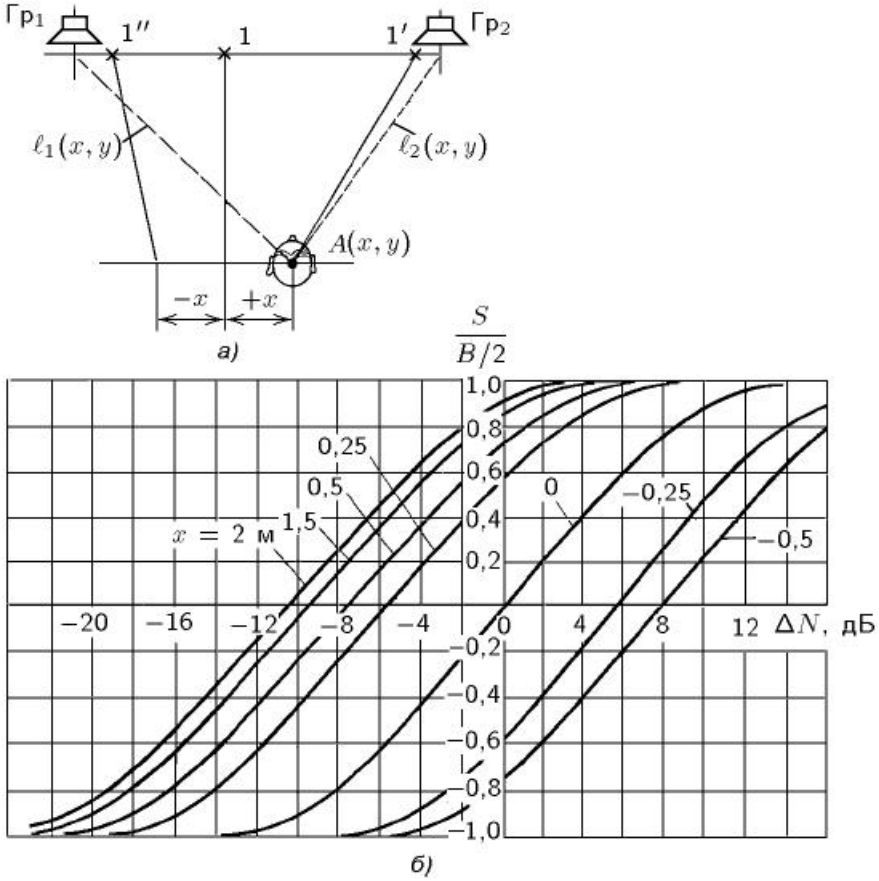


Рис. 1.68. Пример бокового смещения слушателя (*а*) и зависимости относительного смещения КИЗ от бокового смещения слушателя при $\Delta\tau = 0$, $B = 2{,}8$ м, $y = 2$м (*б*):
1 – $x = 2$ м; 2 – $x = 1{,}5$ м; 3 – $x = 0{,}5$ м; 4 – $x = 0{,}25$ м; 5 – $x = 0$;
6 – $x = -0{,}25$ м; 7 – $x = -0{,}5$ м

величин $\Delta\tau_{x,y}$, $\Delta N_{x,y}$ (там, где это еще не приводит к распаду КИЗ). Неизменность форм кривых на рис. 1.68,*б* свидетельствует о независимости действия на орган слуха временных и интенсивностных различий и позволяет характеризовать местоположение каждой кривой величиной ΔN_0, необходимой для возвращения КИЗ в центр базы громкоговорителей. На рис. 1.69,*а* приведены кривые зависимости ΔN_0 от x, а на рис. 1.69,*б* представлена зависимость ΔN_0 от размеров базы громкоговорителей. Момент перехода КИЗ через центр базы громкоговорителей характеризуется наибольшим разбросом экспертопоказаний. Штриховой линией показаны аппроксимированные значения ΔN_0.

Коэффициент эквивалентности. Величины ΔN и $\Delta\tau$ эквивалентны по действию их на орган слуха. Определенные пары значений $\Delta N'$ и $\Delta\tau'$ могут

вызывать одно и то же смещение КИЗ от своего первоначального положения. Их отношение называют коэффициентом эквивалентности:

$$K(x) = \Delta N'(x, y) / \Delta\tau'(x, y),$$

где $\Delta N'(x,y)$ и $\Delta\tau'(x,y)$ — соответственно интенсивностное и временное различия сигналов громкоговорителей, необходимые для возвращения КИЗ в центр базы при расположении слушателя в точке с координатами x и y.

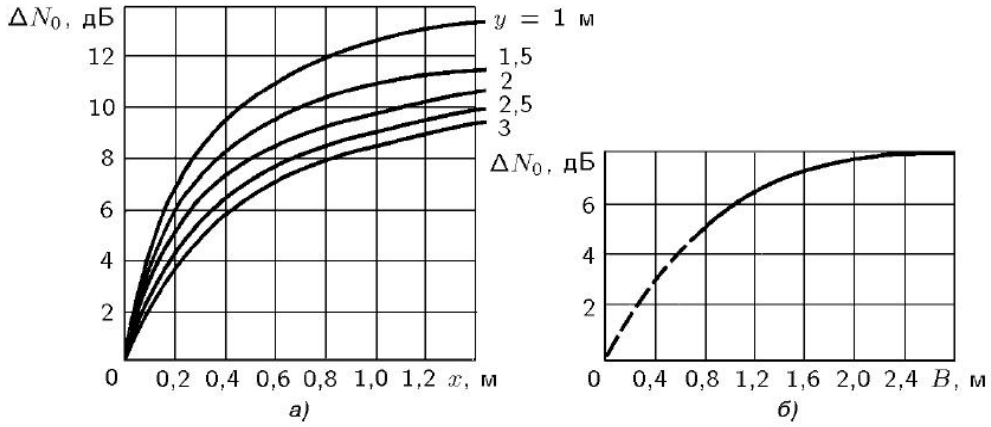


Рис. 1.69.Влияние бокового смещения *x* на величину компенсирующей интенсивностной разности: *а* – при различных *y* и *B* = 2,8 м (1 – *y* = 1 м; 2 – *y* = 1,5 м; 3 – *y* = 2 м; 4 – *y* = 2,5 м; 5 – *y* = 3 м); *б* – при различных значениях базы громкоговорителей (*x* = 0,5 м; *y* = 2 м)

Значение *K*(*x*) не зависит от размеров базы *B* громкоговорителей, расстояния до нее, если y > 1,0 м, и составляет около 10 дБ/мс при x = 0; резко уменьшается с увеличением *x*, составляя уже при *x* = 1 м около 1,5 дБ/мс. Экспериментальная зависимость (сплошная линия на рис. 1.70,а) хорошо

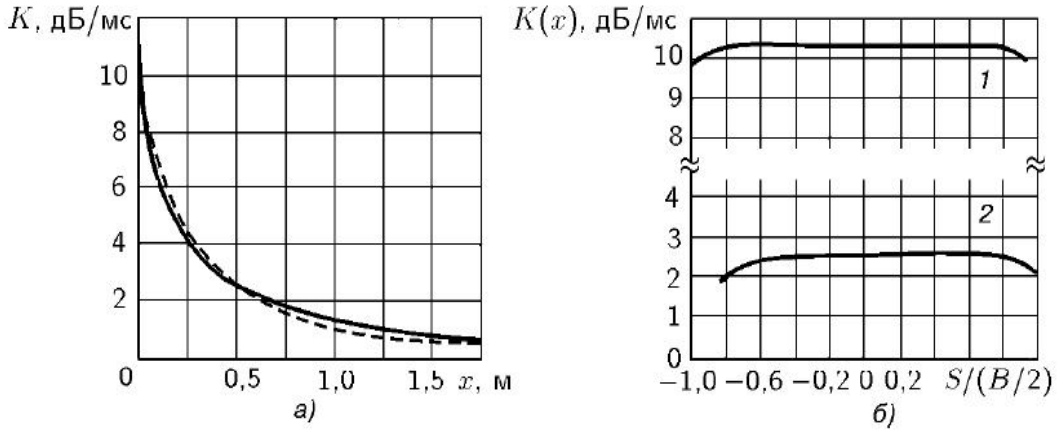


Рис. 1.70. Коэффициент эквивалентности в функции: *а* – бокового смещения слушателя (штриховая линия – аппроксимация выражением (2.5)); *б* – положения КИЗ на линии базы громкоговорителей при симметричном (1) и асимметричном (2) положениях слушателя (*x* = 0,5 м; *y* = 2 м; B = 1,8 м)

аппроксимируется выражением

$$K(x)=[2/(x+0{,}2)]-0{,}3.$$

Здесь x выражено в метрах. Величина $K(x)$ постоянна для КИЗ расположенных в средней части стереопанорамы (рис. 1.70,*б*). Для КИЗ, удаленных от центра базы более, чем на 0,8*B/2*, значение $K(x)$ несколько уменьшается.

Используя понятие коэффициента эквивалентности, нетрудно при одновременном действии на орган слуха ΔN, $\Delta\tau$, $\Delta N_{x,y}$, $\Delta\tau_{x,y}$ перейти к чисто интенсивностной

$$\Delta N_э = \Delta N + \Delta N_{x,y} + K(x)(\Delta\tau + \Delta\tau_{x,y})$$

или чисто временной стереофонии

$$\Delta\tau_э = \Delta\tau + \Delta\tau_{x,y} + (\Delta N + \Delta N_{x,y})/K(x)$$

и затем с помощью зависимостей $S/(B/2) = f_1(\Delta N)$ и $S/(B/2) = f_2(\Delta\tau)$, полученных соответственно при $\Delta\tau = 0$ и $x = 0$ или $\Delta N = 0$ и $x = 0$, найти относительное смещение $S/(B/2)$ кажущегося источника звука. Здесь $\Delta N_э$ и $\Delta\tau_э$ — эквивалентные значения разности уровней и времени запаздывания, вызывающие такое же смещение КИЗ от центра базы громкоговорителей, как и совместно действующие величины ΔN, $\Delta N_{x,y}$, $\Delta\tau$, $\Delta\tau_{x,y}$.

1.10.Моделирование механизма локализации кажущихся источников звука

Кажущиеся источники звука (КИЗ) возникают, когда сигналы, излучаемые громкоговорителями, коррелированны и величина коэффициента корреляции превосходит некоторое пороговое значение $R(\Delta\tau) >= 0{,}1\ldots0{,}15$.

Корреляционная модель локализации КИЗ

Из всех известных моделей, пожалуй, наиболее строгой является модель, предложенная Черри-Сайерсом [1.64;1.65]. Ее структурная схема изображена на рис. 1.71. Входные сигналы гребенкой фильтров разделяются на полосы, по ширине соответствующие частотным группам слуха. Для выделенных пар полосных сигналов $x_1'(t), x_1''(t); x_2'(t), x_2''(t), \ldots, x_m'(t), x_m''(t)$ вычис-

ляются функции взаимной корреляции. В диапазоне частот выше 1,6 кГц полосные сигналы предварительно детектируются и усредняются с целью выделения огибающей.

Все обработанные таким образом пары полосных сигналов затем вводятся в блок распознавания, который определяет, какому из хранящихся в слуховой памяти образцов соответствует полученная совокупность взаимно корреляционных функций.

Положение максимума взаимной корреляционной функции связано с боковым смещением (латерализацией) источника звука. Слияние возбуждений в слуховом центре головного мозга слушателя и образование КИЗ

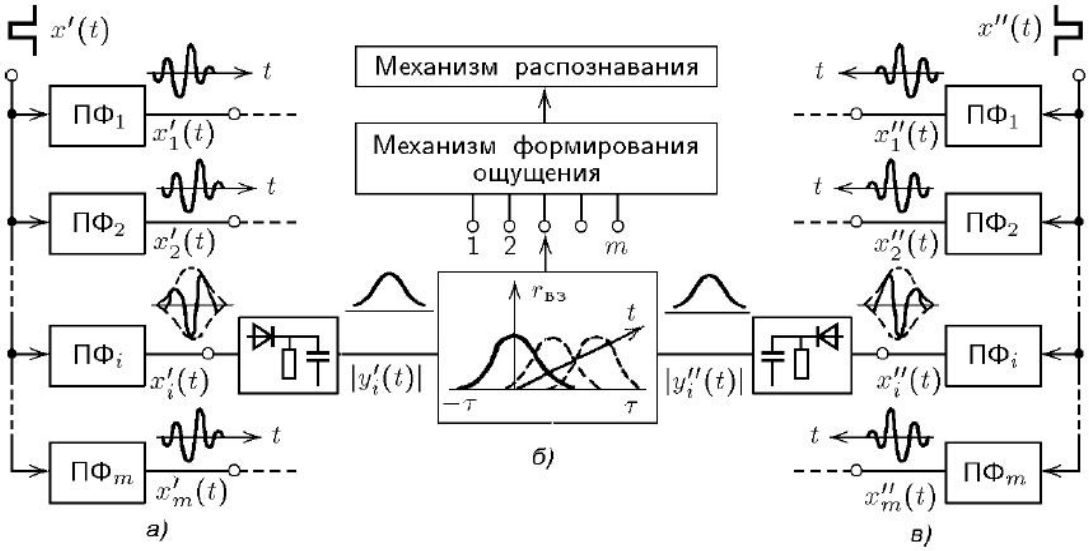


Рис. 1.71. Структурная схема корреляционной модели пространственного слуха: *а*) левое ухо; *б*) центральная нервная система; *в*) правое ухо

становятся возможными, если величины коэффициентов корреляции сигналов, воспринимаемых от разных источников звука, превышают некоторое пороговое значение.

Объясняя процесс образования КИЗ, его латерализацию при введении ΔN и $\Delta\tau$, модель не позволяет рассчитать местоположение этого КИЗ на линии базы громкоговорителей. Однако этот недостаток может быть устранен, если предположить, что:

-оценка временного сдвига бинауральной пары сигналов определяется по величине $\Delta\tau$, при которой функция взаимной корреляции воздействующих сигналов достигает своего максимального значения; в результате такого временного «сканирования» временной сдвиг входных сигналов компенсируется в слуховой системе, когда функция взаимной корреляции достигает своего максимального значения;

-вследствие латерального торможения временной сдвиг сигналов преобразуется в эквивалентную разность уровней путем ослабления по интенсивности запаздывающих сигналов;

-направление на источник звука совпадает с положением максимума функции взаимной корреляции бинауральных сигналов в субъективном слуховом пространстве;

-вводимые в слуховой системе значения Δτ изменяются в точном соответствии с поворотом головы (так называемое «сканирование по азимуту»);

-положение максимального значения функции взаимной корреляции однозначно связано с разностью уровней и временным сдвигом бинауральной пары сигналов.

Покажем на примере двухканальной стереофонии, что введение этих уточнений достаточно для оценки азимута кажущегося источника звука.

Функция локализации

Пусть (рис. 1.72) громкоговоритель $Гр_1$ излучает сигнал *ax(t)*, а громкоговоритель $Гр_2$ – сигнал *aqx(t-Δτ)*, отличающийся от него по ампли амплитуде в *q* раз и запаздывающий по времени на Δτ. Будем считать, что направление на КИЗ в этой модели совпадает с угловым положением максимума функции взаимной корреляции $r_{вз}$ бинауральной пары сигналов,

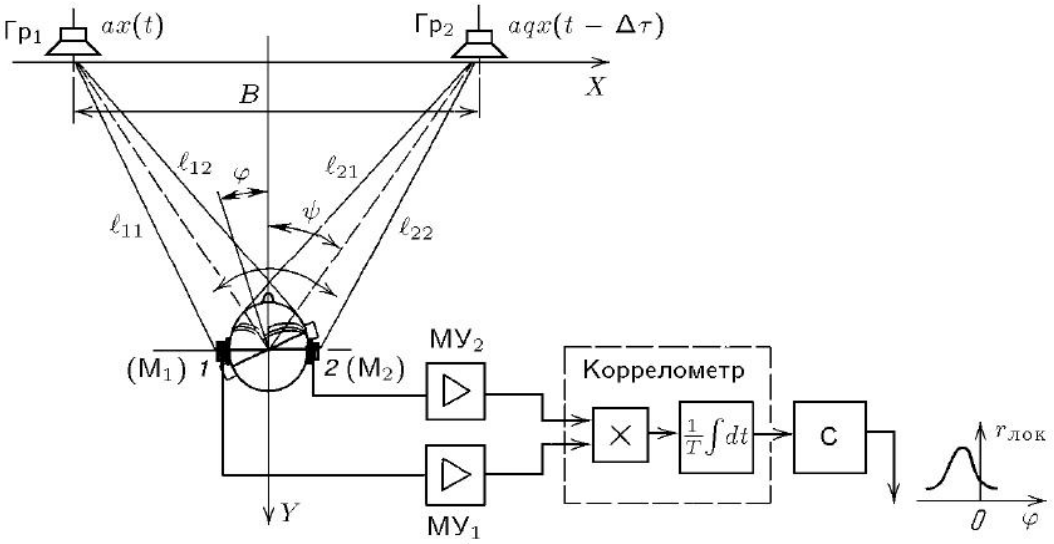


Рис. 1.72. К определению функции локализации $r_{лок}$

воспринятых микрофонами M_1 и M_2:

$$r_{вз}(\varphi) = r_1(\Delta\tau_{12,11}) + r_2(\Delta\tau_{11,22} - \Delta\tau) + \\ + r_3(\Delta\tau_{21,12} + \Delta\tau) + r_4(\Delta\tau_{21,22}) = r_{лок} \quad . \qquad (1.1)$$

Она является суммой четырех корреляционных функций. В этом выражении первое слагаемое $r_1(\Delta\tau_{12,11})$ характеризует воздействие сигнала ле-

вого громкоговорителя на левое 1 и правое 2 уши слушателя; четвертое слагаемое $r_4(\Delta\tau_{21,22})$ то же самое, но для сигнала правого громкоговорителя $Гр_2$. Второе слагаемое $r_2(\Delta\tau_{11,22} - \Delta\tau)$ — результат воздействия сигнала $Гр_1$ на левое ухо 1 и сигнала $Гр_2$ — на правое ухо 2. Третье слагаемое $r_3(\Delta\tau_{21,12} + \Delta\tau)$ - результат перекрестного воздействия сигналов $Гр_1$ и $Гр_2$ соответственно на правое 2 и левое 1 уши слушателя. В данном выражении $\Delta\tau_{12,11}$; $\Delta\tau_{11,22} - \Delta\tau$; $\Delta\tau_{21,12} + \Delta\tau$; $\Delta\tau_{21,22}$ — временные разности соответствующих пар бинауральных сигналов поступающих от $Гр_1$ и $Гр_2$ на левое 1 и правое 2 уши слушателя.

Функция взаимной корреляции $r_{âç}(\varphi)$ сигналов $y_1(t)$ и $y_2(t)$ может быть измерена с помощью корреляционного пеленгатора (рис. 1.72). Он содержит искусственную голову с микрофонами $М_1$ и $М_2$, микрофонные усилители $МУ_1$ и $МУ_2$, перемножитель сигналов X, интегратор $\frac{1}{T}\int dt$ и самописец С, фиксирующий измеренные значения на диаграммной ленте.

При повороте искусственной головы (рис. 1.73,*а*) будут меняться величины $l_{11}, l_{12}, l_{21}, l_{22}$, а следовательно, и соответствующие им временные задержки $\tau_{11} = (l_{11}/c),\ \tau_{12} = (l_{12}/c), \tau_{21} = (l_{21}/c), \tau_{22} = (l_{22}/c)$ и значения $\Delta\tau_{12,11} = \tau_{12} - \tau_{11},\ \Delta\tau_{12,21} = \tau_{12} - \tau_{21},\ \Delta\tau_{21,22} = \tau_{21} - \tau_{22},\ \Delta\tau_{11,22} = \tau_{11} - \tau_{22}$. Это, в свою очередь, вызовет изменение функций, составляющих $r_{вз}(\varphi)$. Зависимости $\Delta\tau_{12,11}$, $\Delta\tau_{11,22}$, $\Delta\tau_{21,12}$, $\Delta\tau_{21,22}$ в функции от угла поворота φ легко вычислить теоретически, представив упрощенно искусственную голову в форме шара диаметром, как это обычно принято, равным $D' = 16{,}6$ см. Вид этих кривых показан на рис. 1.73,*б*. Здесь по оси абсцисс отложены значения угла поворота искусственно головы относительно медианной плоскости, в град.,

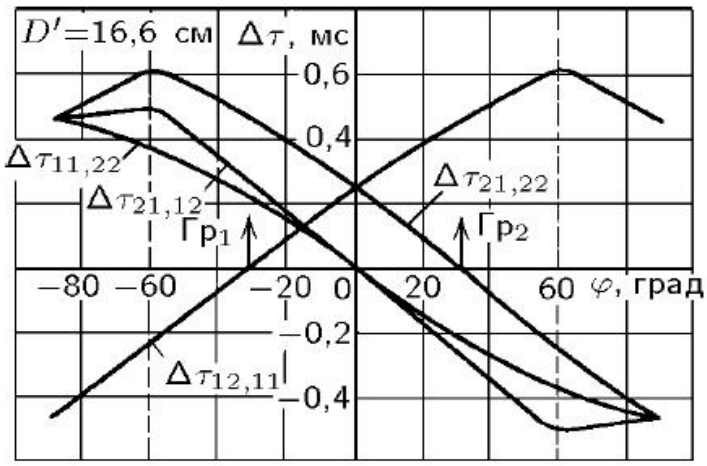


Рис. 1.73. Зависимость бинауральных временных сдвигов $\Delta\tau_{11,12}$, $\Delta\tau_{21,22}$, $\Delta\tau_{11,22}$, $\Delta\tau_{21,12}$ от угла поворота искусственной головы, аппроксимированной шаром

по оси ординат – значения временных сдвигов, в мс. Заметим, что функцию $r_1(\Delta\tau_{12,11})$ можно измерить отдельно от всей суммы, выключив правый громкоговоритель, а функцию $r_4(\Delta\tau_{21,22})$ – выключив левый громкоговоритель.

Зависимость $r_{вз} = f(\varphi)$ называют функцией локализации $r_{лок}$. Если сигналы, излучаемые Гр$_1$ и Гр$_2$, имеют вид $aqx(t - \Delta\tau')$ и $ax(t)$ и представляют собой белый шум в полосе частот от ω_1 до ω_2, то их функция корреляции

$$r(\Delta\tau') = \frac{a^2 q\Delta\omega}{2}\frac{\sin(\Delta\omega\Delta\tau'/2)}{\Delta\omega\Delta\tau'/2}\cos(\omega_0\Delta\tau'), \qquad (1.2)$$

где $\Delta\omega = \omega_2 - \omega_1$ — полоса круговых частот; $\omega_0 = (\omega_1+\omega_2)/2$ — средняя круговая частота. Расчетные выражения для вычисления составляющих $r_1(\Delta\tau_{12,11})$; $r_2(\Delta\tau_{11,22} - \Delta\tau)$; $r_3(\Delta\tau_{21,12} + \Delta\tau)$; $r_4(\Delta\tau_{21,22})$, аналогичны выражению (1.2). Разница состоит лишь в том, что для каждой из них существует свое максимальное значение, равное $\alpha^2\Delta\omega/2$ для $r_1(\Delta\tau_{12,11})$; $\alpha^2 q^2\Delta\omega/2$ для $r_4(\Delta\tau_{21,22})$; $\alpha^2 q\Delta\omega/2$ для $r_2(\Delta\tau_{11,22}) - \Delta\tau)$ и $r_3(\Delta\tau_{21,12} + \Delta\tau)$, а также свой временной сдвиг $\Delta\tau'$, указанный для них в круглых скобках.

Поведение функции локализации при разных ситуациях (интенсивностная, временная и смешанная стереофония) показано на рис. 1.74. Однако, угловое положение максимального значения функции локализации совпадает с направлением на КИЗ только в случае чисто интенсивностной стереофонии, когда $\Delta\tau = 0$ и входное поворотное устройство корреляционного пеленгатора расположено на оси симметрии Y громкоговорителей Гр$_1$ и Гр$_2$ системы воспроизведения (рис. 1.74,*а* и *б*). Расчеты показывают, что при расположении входного поворотного устройства на оси симметрии громкоговорителей Гр$_1$ и Гр$_2$ и $\Delta\tau \neq 0$ (временная стереофония) имеет место (рис. 1.74,*в*) смещение максимумов функций $r_2(\Delta\tau_{11,22} - \Delta\tau)$ и $r_3(\Delta\tau_{21,12} + \Delta\tau)$ в разные стороны от значения $\varphi = 0$. При этом функция локализации уплощается, ее максимум не изменяет своего углового положения, но становится весьма неопределенным в то время, как слушатель уверенно локализует КИЗ и отмечает его смещение в сторону громкоговорителя, излучающего опережающий сигнал. Не обеспечивается также получение правильных результатов и при асимметричном положении входного поворотного устройства пеленгатора вследствие появляющегося при этом временного сдвига $\Delta\tau_{x,y}$. Последний обусловлен различием расстояний до громкоговорителей Гр$_1$ и Гр$_2$.

Для преодоления этих трудностей необходима предварительная коррекция входных сигналов, выполняемая так же, как об этом свидетельствуют результаты новейших исследований, и в слуховой системе человека. Разница по времени $\Delta\tau$ заменяется эквивалентной разностью уровней $\Delta N_{экв} = K_x\Delta\tau$ с использованием уже введенного ранее коэффициента эквивалентности K_x. Вполне понятно, что если эта коррекция будет выполняться для

сигналов, воспринятых микрофонами M_1 и M_2 (рис. 1.72), то количественные соотношения между величинами $\Delta N_{ЭКВ}$ и $\Delta\tau$ будут иными по сравнению с тем, что имело бы место, если бы она проводилась для сигналов, подводимых к $Гр_1$ и $Гр_2$.

Введение величины $\Delta\tau_{ОПТ}$, компенсирующей временной сдвиг $\Delta\tau$ сигналов, «возвращает» функции $r_2(\Delta\tau_{11,22} - \Delta\tau)$ и $r_3(\Delta\tau_{21,12} + \Delta\tau)$ в свое первоначальное положение, когда максимальные значения последних

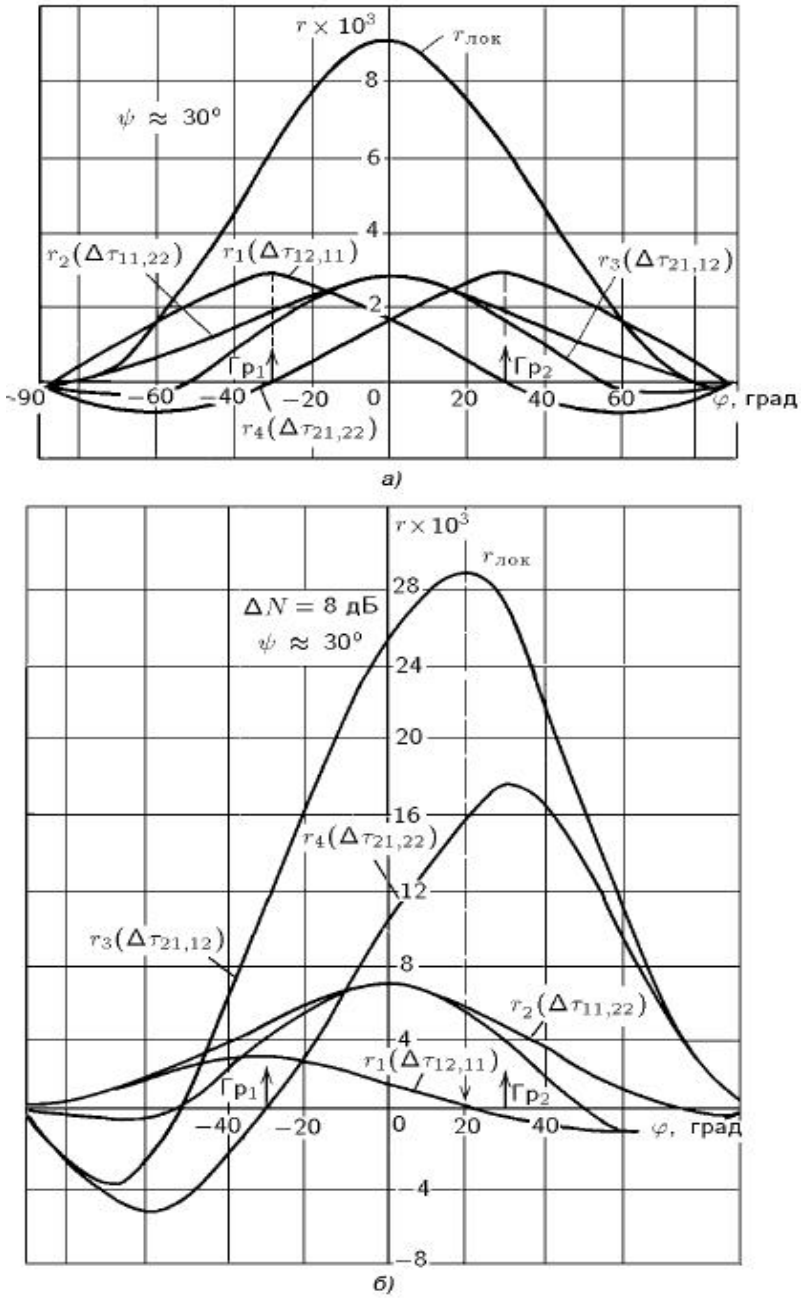


Рис. 1.74, *а* и *б*. Функция локализации и ее составляющие (белый шум в полосе частот 100…1000 Гц; B = 2,8 м; $y = B$): *а* – при воспроизведении тождественных сигналов; *б* – при интенсивностной стереофонии (x = 0)

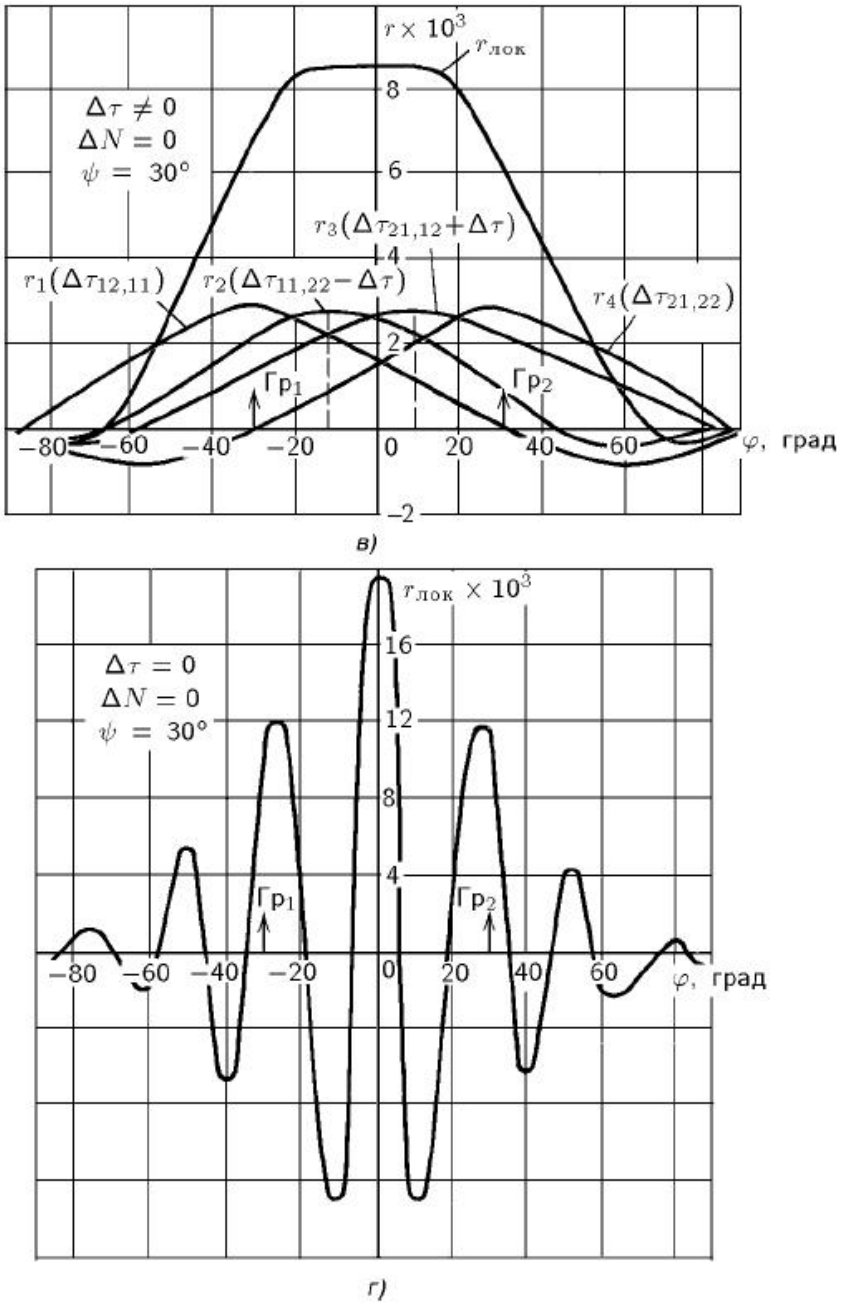


Рис. 1.74, *в* и *г*. Функция локализации и ее составляющие (белый шум в полосе частот 100…1000 Гц; *B* = 2,8 м; *y* = *B*): *в* – при временной стереофонии (*x* = 0); *г* – многозначность функции локализации при $F_0 > 1000$ Гц (белый шум в полосе частот 4000…6000 Гц)

совпадают с направлением на центр базы. Ослабление же запаздывающего сигнала по интенсивности приводит случай чисто временной стереофонии к чисто интенсивностной стереофонии.

И еще одно обстоятельство непременно должно быть учтено. При достаточно высокой средней частоте (F_0 > 1000 Гц) полосы шума каждое из слагаемых функции локализации $r_{лок}$, вычисленное с помощью выражения

(1.2), становится многозначным ввиду быстрого изменения сомножителя $\cos(\omega_o\Delta\tau')$. Вследствие этого функция локализации также становится многозначной (рис. 1.74,*г*). При этом ее главный максимум в общем случае уже не совпадает с направлением на КИЗ. Для устранения этого затруднения уместно вспомнить следующее: механизм локализации, являясь инерционным элементом органа слуха (время адаптации слуха на изменение направления составляет около 120…150 мс), реагирует не на мгновенные значения звукового сигнала, а на его огибающую. Последняя получается путем выпрямления (линейный детектор) и усреднения (*RC*-фильтр) последних. Напомним, что длительность слуховой памяти составляет 30…50 мс. Если допустить, что в слуховой системе происходит выделение огибающей сигнала, то выражение (1.2) для расчета слагаемых $r_{лок}$ при $F_0 >$ 1000...1500 Гц преобразуется к виду

$$r^{'}(\Delta\tau') = a^2 q \frac{\Delta\omega}{2}\left|\frac{\sin(\Delta\omega\Delta\tau'/2)}{\Delta\omega\Delta\tau'/2}\right|.$$

Учет этого обстоятельства и замена $r(\Delta\tau')$ на $r'(\Delta\tau')$ приводит к тому, что и в этом случае функция локализации имеет один четкий максимум, а угловое положение последнего совпадает с направлением на КИЗ.

Сопоставление теоретических результатов и данных экспертиз показывает, что учет всех изложенных выше дополнений приводит к тому, что функция локализации имеет один четкий максимум. При этом его угловое положение при любых условиях проведения эксперимента всегда совпадает с направлением на кажущийся источник звука. Таким образом, корреляционный пеленгатор с блоками дополнительной коррекции представляет собой прибор для оценки азимута КИЗ при двухканальном стереовоспроизведении. Эти дополнения показаны на рис. 1.75. Измерительные устройства содержат: «искусственную голову с микрофонами $М_1$ и $М_2$, микрофонные усилители $МУ_1$ и $МУ_2$, устройства аналогичные электрической части измерителя уровней: двухполупериодные выпрямители ($В_1$ и $В_2$,) и сглаживающие фильтры ($Ф_1$ и $Ф_2$,), собственно коррелометр (КА), содержащий перемножитель сигналов (х) и интегратор $\frac{1}{T}\int dt$, и регистрирующий прибор (самописец С или стрелочный индикатор – И). Кроме того, необходимой частью корреляционного пеленгатора является блок предварительной коррекции сигналов (БКС). С его помощью при $\Delta\tau\neq0$ или асимметричном положении входного поворотного устройства относительно $Гр_1$ и $Гр_2$ перед началом измерений выполняется предварительная коррекция сигналов (либо до громкоговорителей, как это показано на рис. 1.75,*а*, либо после микрофонов – рис. 1.75,*б*.

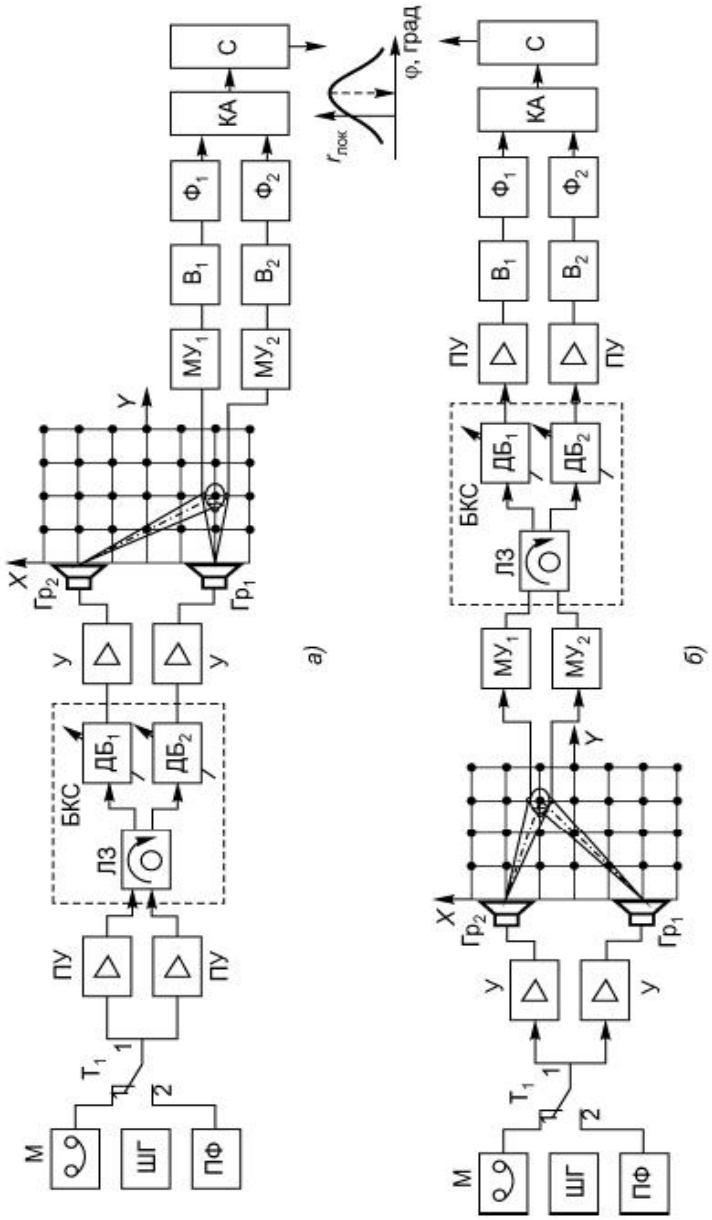


Рис. 1.75.Структурные схемы измерителей направления на КИЗ при двухканальной стереофонии: а - БКС в передающей части устройства; б – то же самое, но в приемной части устройства

При проведении измерений заглушенной камеры не требуется. Это несомненное достоинство данного метода. Как показали исследования, достаточно хорошее совпадение результатов измерений и экспертиз на частотах ниже 1000 Гц достигается и без применения искусственной головы. Здесь достаточно оба микрофона установить на штативе. Причем увеличение базы микрофонов (расстояния между ними) обостряет «резонансный» характер изменения величины $r_{лок}$ в зависимости от φ и, следовательно, повышает точность измерений.

В случае многоканальной стереофонии картина существенно усложняется, но об этом будет сказано позже.

Ассоциативная модель слуха и оценка азимута источника звука

Ассоциативная модель локализации источника звука [1.57; 1.64] предполагает наличие двух последовательных этапов переработки информации в слуховой системе: ассоциации места действительного источника звука в пространстве и ассоциации формы, где возможны образование КИЗ и оценка его азимута.

Звуковая волна, распространяясь от источника звука (рис. 1.76,*а*) соответственно к левому и правому входам слухового анализатора, претерпевает изменения, вызванные частотно-зависимым затуханием звука в воздухе с

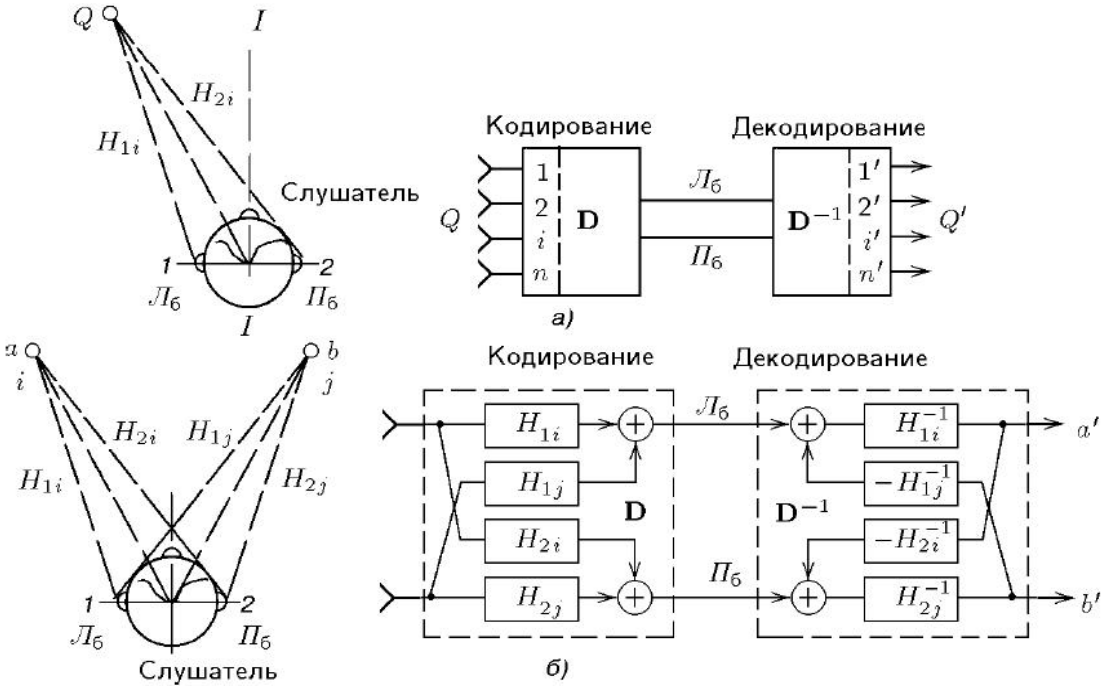


Рис. 1.76. К пространственному кодированию и декодированию одного (*а*) и двух (*б*) действительных источников звука

расстоянием, дифракционными явлениями, определяемыми формой головы и ушных раковин слушателя. Все эти изменения могут быть однозначно описаны парой образующих матрицу **D** передаточных функций H_{1i} и H_{2i} линейной цепи, расположенной между источником звука и левым и правым входами слуховой системы.

Матрица **D** однозначно определяет место действительного источника звука в пространстве. Этот процесс можно представить как пространственное кодирование источника звука. При этом бинауральная пара сигналов $Л_б$

и $П_б$, соответствующая источнику звука, расположенному в точке i (рис. 1.76,*а*),

$$Л_б = H_{1i}Q\ ; \quad П_б = H_{2i}Q\ ,$$

где Q – сигнал, излучаемый источником звука, H_{1i} и H_{2i} – коэффициенты передачи, описывающие все те изменения, которые претерпевает звуковая волна, распространяясь от места i нахождения источника звука до левого 1 и правого 2 уха слушателя.

При наличии двух действительных источников звука, расположенных в местах i и j и излучающих соответственно сигналы a и b (рис. 1.76, *б*), результат их пространственного кодирования можно представить в виде

$$Л_б = H_{1i}\,a + H_{1j}b; \qquad П_б = H_{2j}\,b + H_{2i}\,a.$$

Здесь H_{1i} , H_{1j} , H_{2j} , H_{2i} –зависящие от места коэффициенты передачи, описывающие изменения, претерпеваемые звуковой волной при распространении от каждого источника звука i и j соответственно до левого 1 и правого 2 уха слушателя С.

Пространственное декодирование заключается в разделении (селекции) пар бинауральных сигналов по принципу места. Этот этап обработки информации в слуховой системе является первым и носит название ассоциации места. В памяти слуха для каждой совокупности мест $\{i,j\}$ существует инверсная матрица $\mathbf{D}^{-1}$, коэффициенты передачи которой для каждого из пары бинауральных сигналов обратны соответствующим коэффициентам матрицы **D**. С ее помощью осуществляется разделение (селекция) сигналов источников звука по принципу места (рис. 1.76,*б*). Для источника звука, расположенного в месте i (см. рис. 1.76,*а*), матрица декодирования $\mathbf{D}^{-1}$ адаптивным путем принимает коэффициенты передачи равные $1/H_{1i}$ и $1/H_{2i}$, что обеспечивает выделение сигнала Q'. Действительно,

$$Л_б \cdot 1/H_{1i} = Q' \qquad \text{и} \qquad П_б \cdot 1/H_{2i} = Q'.$$

При наличии двух действительных источников звука i и j коэффициенты передачи инверсной матрицы $\mathbf{D}^{-1}$, равные

$$\frac{H_{2j}Л_б - H_{1i}П_б}{H_{1i}H_{2j} - H_{2i}H_{1i}} \quad и \quad \frac{H_{1j}П_б - H_{2i}Л_б}{H_{1i}H_{2j} - H_{2i}H_{1i}} \quad ,$$

обеспечивают распознавание сигналов a' и b', отличающихся от исходных a и b на величину погрешности. Процесс декодирования рассматривается как следствие ассоциации признаков бинауральных пар сигналов источни-

ков, подвергнутых пространственному кодированию, с образцами, хранящимися в слуховой памяти.

После пояснения процессов пространственного кодирования и декодирования сигналов источников звука можно перейти к изложению общих принципов функционирования ассоциативной модели (рис. 1.77). Предварительно заметим, что ассоциативный метод обработки информации, по-видимому, свойственен всем «живым» системам с памятью.

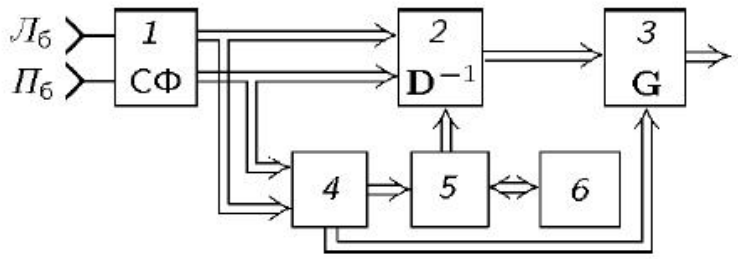


Рис. 1.77. Ассоциативная модель пространственного слуха: 1- система полосовых фильтров, тождественных по ширине критическим полосам слуха; 2 – адаптивный фильтр, описываемой матрицей $\mathbf{D}^{-1}$; 3 – этап ассоциации формы **G** ; 4 – оценка корреляции бинауральной пары сигналов; 5 – блок идентификации корреляционных образцов; 6 – слуховая память

Бинауральные сигналы $Л_б$ и $П_б$ источников звука, подвергнутые пространственному кодированию, в периферийном отделе слуховой системы разделяются системой фильтров СФ на полосы частот приблизительно одинаковой относительной ширины, называемые критическими полосами (или частотными группами) слуха. Дальнейшая обработка этих выделенных пар полосных сигналов осуществляется в центральной части слуховой системы раздельно в два этапа.

На этапе ассоциации места сигналы источников звука отделяются друг от друга путем их пространственного декодирования. Процесс переработки информации на этом этапе можно описать действием адаптивного фильтра $\mathbf{D}^{-1}$, параметры которого регулируются на основе ассоциативного распознавания образов. Путем сравнения признаков, полученных в результате пространственного кодирования сигналов источников звука, с приобретенными на основании жизненного опыта эталонными образцами, хранящимися в слуховой памяти, сигналы источника распознаются слушателем, и адаптивный фильтр $\mathbf{D}^{-1}$ принимает коэффициенты передачи, обратные матрице **D**. По-видимому, целесообразным сигналом для ассоциативного процесса распознавания является бинауральный корреляционный образец.

На выходе адаптивного фильтра $\mathbf{D}^{-1}$ сигнал источника «освобождается» от всех тех изменений, которые были внесены на этапе пространственного кодирования.

Таким образом осуществляется ассоциативная селекция источников звука, определяющая их место в пространстве. Информация о месте источника звука и соответствующий ему сигнал, освобожденный от искажений, внесенных на этапе пространственного кодирования, передаются дальше.

На этапе ассоциации формы **G** представлены все механизмы слуха, предназначенные для анализа разделенных по принципу места сигналов действительных источников звука. Сюда относятся механизмы слияния возбуждений и образования КИЗ, анализа тембра, динамики звука, высоты тона, уровня громкости и т.д. Распознавание звуковых образов на этапе ассоциации формы есть также и результат обращения к слуховой памяти, представляющей собой банк данных, где хранятся соответствующие эталонные образцы, приобретенные на основании жизненного опыта. Нельзя узнать и идентифицировать звучание, если человек его ни разу не слышал. Так же как этап ассоциации места содержит механизм селекции локализованных возбуждений, вызванных сигналами отдельных источников звука, этап ассоциации формы содержит механизм селекции последних по форме.

В соответствии с ассоциативной моделью слуха проблема локализации звуковых образов в пространстве трактуется следующим образом: одиночный действительный источник звука всегда вызывает одну ассоциацию места, которая и определяет его местоположение в пространстве; два пространственно разнесенных источника звука на этапе ассоциации места также разделены друг от друга. Если сигналы этих источников не коррелированны, то на этапе распознавания формы они вызывают две ассоциации формы, воспринимаются как два раздельных звуковых образа, положения в пространстве которых по-прежнему определяются по принципу локализации места на первом этапе обработки информации. Наличие корреляции между сигналами источников звука на этапе ассоциации формы приводит к слиянию событий слушания и образованию одного кажущегося источника звука. В этом случае локализация КИЗ является уже следствием двух этапов переработки информации – ассоциации места и ассоциации формы.

При воспроизведении сигнал $y_j(t)$ каждого громкоговорителя СВ кодируется пространственным фильтром, в качестве которого выступают голова и ушные раковины слушателя. Процесс пространственного кодирования сигналов громкоговорителей записывается в виде

$$Л_б = \sum_{j=1}^{N'} H_{1j} y_j(t); \quad П_б = \sum_{j=1}^{N'} H_{2j} y_j(t),$$

где $Л_б$ и $П_б$ — левый и правый бинауральные слуховые сигналы; N' – число каналов воспроизведения адаптивного декодирующего устройства (АДУ) или громкоговорителей системы воспроизведения (СВ), H_{1j} и H_{2J} – коэффициенты передачи, описывающие изменения, которые претерпевает звуковая

волна, распространяясь от j-го громкоговорителя СВ к левому 1 (H_{1j}) и правому 2 (H_{2j}) ушам слушателя. Множество $\{H_{1j},H_{2j}\}_N$ образует матрицу пространственного кодирования **D** сигналов действительных источников звука – громкоговорителей СВ.

1.11.Ассоциативная модель слуха и передача пространственной информации в звуковых системах повышенного качества звучания

Рассмотрим применение ассоциативной модели слуха для оценки азимута КИЗ на примере стереофонической системы повышенного качества звучания с панорамным кодированием сигналов источников звука (рис. 1.78). Здесь: 1, 2, ..., N — каналы первичных звуковых сигналов (сигналов источников звука), каждый такой сигнал $x_i(t)$ формирует на стороне

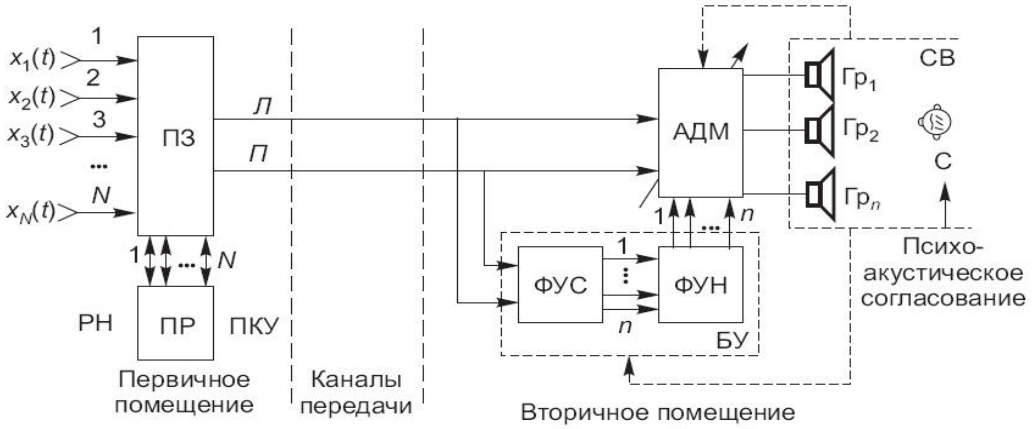


Рис. 1.78.Структурная схема двухканальной стереофонической системы повышенного качества звучания: ПЗ – пульт звукорежиссера; ПКУ – панорамно-кодирующее устройство (РН – регулятор направления, ПР – панорамный регулятор); АДМ – адаптивная декодирующая матрица; ФУС – формирователь управляющих сигналов; ФУН – формирователь управляющих напряжений; БУ – блок управления;
СВ – система воспроизведения; С - слушатель

воспроизведения свой КИЗ; ПКУ — панорамно-кодирующее устройство, которое входит в состав пульта звукорежиссера, с его помощью сигналы источников звука без промежуточных преобразований непосредственно преобразуются в двухканальный сигнал Л(t) и П(t); АДМ – адаптивное декодирующая матрица; $Гр_1$, $Гр_2$,…,$Гр_n$ – громкоговорители системы воспроизведения СВ; С – слушатель; ФУС – формирователь управляющих сигналов; ФУН – формирователь управляющих напряжений, обеспечивающий требуемый закон изменения и глубину регулировки коэффициентов передачи управляемой АДМ, а также время установления и восстановления ее коэффициентов передачи.

С учетом ассоциативной модели слуха в любой стереофонической системе процессы кодирования, передачи, декодирования, воспроизведения и восприятия пространственной информации (рис. 1.79) можно представить выражением:

$$\hat{x}_i(t,\varphi) = \mathbf{ABDCG}x_i(t), \qquad (1.3)$$

где $\hat{x}_i(t,\varphi)$ — оценка слухового сигнала; каждый такой сигнал формирует

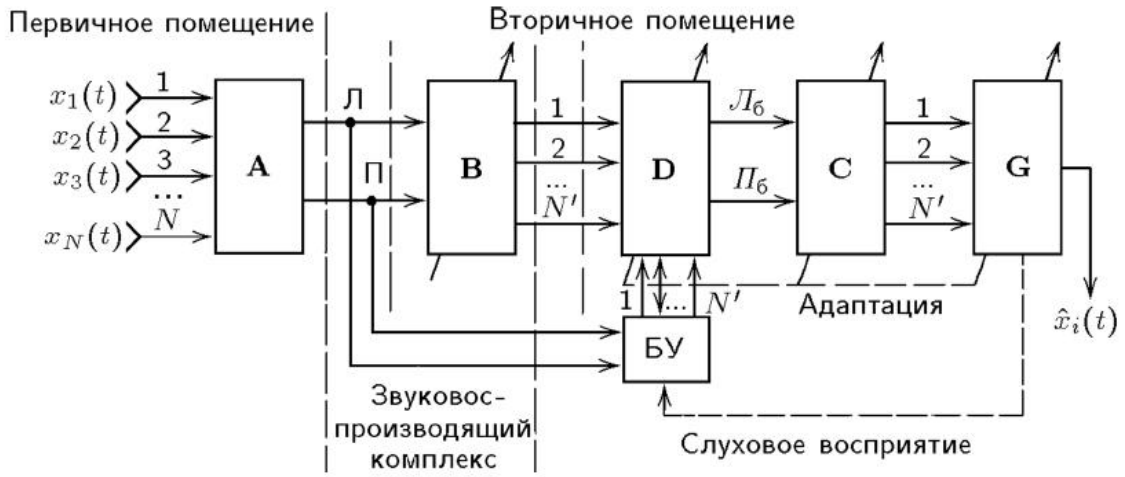


Рис. 1.79. Кодирование, передача, воспроизведение и восприятие пространственной информации

при слуховом восприятии в жилом помещении i - й кажущийся источник звука; t – текущее время; φ – азимутальный угол этого кажущегося источника звука; **A** – матрица панорамного кодирования множества сигналов источников звука $\{x_i(t)\}_N$ в левый Л(t) и правый П(t) сигналы стереопары. Уравнения кодирования при этом имеют вид:

$$Л(t) = \sum_{i=1}^{N} a_{1i} x_i(t); \;\; П(t) = \sum_{i=1}^{N} a_{2i} x_i(t) \;,$$

где $x_i(t)$ – сигнал i-го источника звука. Каждый такой сигнал на стороне воспроизведения образует свой кажущийся источник звука; множество сигналов $\{x_i(t)\}_N$ формирует стереопанораму в помещении прослушивания; a_{1i} и a_{2i} – пары коэффициентов панорамного кодирования сигналов каждого источника звука, значения этих коэффициентов зависят от угла локализации образуемого сигналами каждой из этих пар кажущегося источника звука; φ_i – угол, под которым этот кажущийся источник звука локализуется слушателем при воспроизведении этой пары сигналов. Необходимо отметить также, что

$$\Delta N_i = 20lg(a_{2i}/a_{1i}); \qquad a^2_{2i} + a^2_{1i} = 1.$$

Здесь ΔN_i – разность уровней, определяющая оценку азимута i-го КИЗ, дБ; **B** – матрица панорамного декодирования сигналов Л(t) и П(t), с помощью которой они преобразуются к виду

$$y_j\,(t) = b_{j1}Л(t) + b_{j2}П(t),$$

где $y_j(t)$ — сигнал, воспроизводимый j-м громкоговорителем СВ в жилом помещении; N' – число громкоговорителей, образующих СВ; b_{j1} и b_{j2} — пары коэффициентов декодирования для сигналов Л(t) и П(t); **D** – матрица пространственного кодирования сигналов громкоговорителей СВ при их слуховом восприятии; **C** - матрица пространственного декодирования сигналов действительных источников звука – громкоговорителей в слуховой системе слушателя; **G** – матрица, характеризующая обработку векторных сигналов громкоговорителей в слуховой системе человека при оценке азимута КИЗ.

При воспроизведении сигнал $y_j(t)$ каждого громкоговорителя кодируется пространственным фильтром, в качестве которого выступают голова и ушные раковины слушателя. Процесс пространственного кодирования сигналов громкоговорителей записывается в виде

$$Л_б = \sum_{j=1}^{N'} H_{1j} y_j(t); \quad П_б = \sum_{j=1}^{N'} H_{2j} y_j(t),$$

где $Л_б$ и $П_б$ – левый и правый бинауральные слуховые сигналы; N' – число каналов воспроизведения адаптивного декодирующего устройства (АДУ) или громкоговорителей СВ, H_{1j} и H_{2J} – коэффициенты передачи, описывающие изменения, которые претерпевает звуковая волна, распространяясь от j-го громкоговорителя СВ к левому 1 (H_{1j}) и правому 2 (H_{2j}) ушам слушателя. Множество $\{H_{1j},H_{2j}\}_N$ образует матрицу пространственного кодирования **D** сигналов действительных источников звука – громкоговорителей СВ.

Рассмотрим подробнее процесс переработки пространственной информации в слуховой системе человека. Допустим, что система воспроизведения состоит из нескольких громкоговорителей. Местоположение i каждого из них в пространстве обозначим цифрами 1,2,3,…,n. Предположим, что на вход системы звукопередачи один первичный сигнал $x_i(t)$, а громкоговорители СВ излучают соответственно сигналы вида $a_1x_i(t)$, $a_2x_i(t)$, $a_3x_i(t)$,….., $a_nx_i(t)$, где $a_1, a_2, a_3,...., a_n$, - амплитуды сигнала; $x_i(t)$ – его временная функция, причем $-1 <= x_i(t) <= +1$.

Итак, процесс пространственного кодирования сигналов громкоговорителей СВ может быть записан следующим образом:

$$Л_б = \sum_{i=1}^{n} H_{1i} a_i x_i(t)\,; \quad П_б = \sum_{i=1}^{n} H_{2i} a_i x_i(t)$$

или в матричной форме:

$$\begin{pmatrix} Л_б \\ П_б \end{pmatrix} = \begin{pmatrix} H_{11}\ H_{12}\ H_{13} \ldots H_{1n} \\ H_{21}\ H_{22}\ H_{23} \ldots H_{2n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ a_n \end{pmatrix}.$$

Здесь H_{1i} и H_{2i} – коэффициенты передачи, описывающие все изменения, которые претерпевает звуковая волна, распространяясь от громкоговорителя i к левому 1 (H_{1i}) и правому 2 (H_{2i}) ушам слушателя; иными словами первая цифра индекса при H соответствует левому 1 или правому 2 входам слухового анализатора, а вторая – порядковому номеру громкоговорителя СВ; $Л_б$ и $П_б$ – левый и правый бинауральные сигналы.

В результате пространственного декодирования сигналов $Л_б$ и $П_б$ (этап ассоциации места) сигналы громкоговорителей $a_i x_i(t)$ должны быть отделены друг от друга. Это значит, что формально должно выполняться равенство

$$\begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_n \end{pmatrix} = C \begin{pmatrix} Л_б \\ П_б \end{pmatrix} = C \begin{pmatrix} H_{11}\ H_{12} \ldots H_{1n} \\ H_{21}\ H_{22} \ldots H_{2n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ a_n \end{pmatrix}.$$

Данное условие может быть выполнено, если произведение матриц **D·C** равно 1, поэтому матрица **C** должна быть обратной по отношению к матрице **D**.

Обозначим $\mathbf{C}=\mathbf{D}^{-1}$ и перепишем с учетом этого выражение (1.3):

$$\hat{x}_i(t,\varphi) = \mathbf{ABDD^{-1}G}x_i(t). \qquad (1.4)$$

По определению элемент d_{ik} обратной матрицы $\mathbf{D}^{-1}$ равен транспонированному алгебраическому дополнению Δ_{kj} соответствующего элемента исходной матрицы **D**, деленному на определитель Δ:

$$d_{ik}=(\Delta_{kj}/\Delta).$$

В уравнении (1.4) умножение матриц выполняется справа налево. Заметим, что обратная матрица $\mathbf{D}^{-1}$ существует, если матрица **D** является квадратной

и ее определитель Δ не равен нулю. Оба эти условия выполняются не всегда. Если матрица **D** не квадратная, то уравнение **C**·**D** = **E** (где **E** – единичная матрица) все-таки можно решить, когда матрица (**D**·**D**$^{'}$) будет квадратной. Если у матрицы (**D**·**D**$^{'}$) есть обратная матрица (**D**·**D**$^{'}$)$^{-1}$ при Δ≠0, то матрица **D**$^{'}$(**D**·**D**$^{'}$)$^{-1}$ является псевдообратной матрицей **D**$^{+}$ матрицы **D**, т.е

$$\mathbf{D}^{+} = \mathbf{D}'(\mathbf{DD}')^{-1},$$

где **D'** – матрица, полученная из **D** транспонированием ее элементов, а (**DD'**)$^{-1}$ – обратная матрица по отношению к квадратной матрице (**DD'**). Размер матрицы (**DD'**) определяется числом громкоговорителей СВ. Отметим, что единственная обратная матрица (**DD'**)$^{-1}$ будет существовать, если число строк в **D** не больше числа столбцов.

Итак, с учетом изложенного выше выражения (1.3 и 1.4) можно записать как

$$\hat{x}(t,\varphi) = A \cdot B \cdot D \cdot D^{-1} \cdot G \cdot x_i(t) \quad при \quad n = 2,$$

$$\hat{x}(t,\varphi) = A \cdot B \cdot D \cdot D^{+} \cdot G \cdot x_i(t) \quad при \quad n > 2.$$

Поясним подробнее процедуру получения псевдообратной матрицы **D**$^{+}$ на примере системы воспроизведения, состоящей из трех громкоговорителей (рис. 1.80,*а*). Для данной системы воспроизведения матрица пространственного кодирования

$$D = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \end{pmatrix},$$

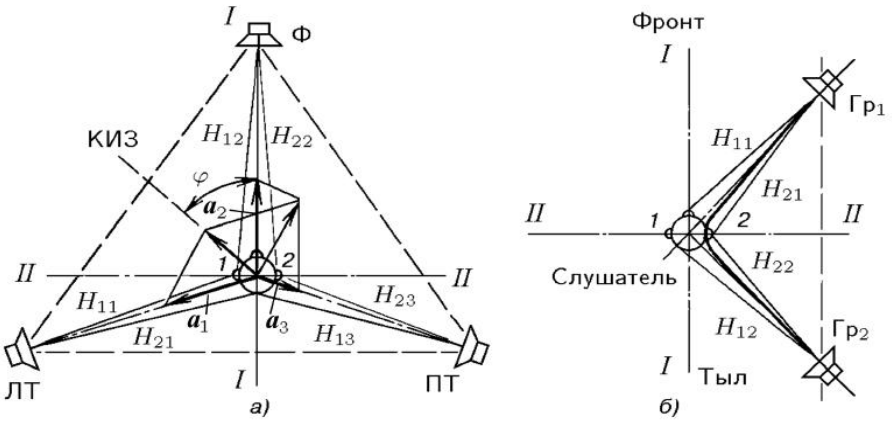


Рис. 1.80. К пояснению локализации КИЗ в векторной модели оценки азимута: *а* – пример оценки азимута КИЗ в системе воспроизведения «треугольник» на этапе ассоциации формы; *б* – случай невозможного слияния звучаний пары громкоговорителей

соответствующая ей транспонированная матрица имеет вид

$$D^{'} = \begin{pmatrix} H_{11} & H_{21} \\ H_{12} & H_{22} \\ H_{13} & H_{23} \end{pmatrix}.$$

Матрица $\mathbf{D}^{'}$ всегда (при любом значении n) соответствует матрице **D**, так как число столбцов первой равно числу строк второй. Поэтому умножение матриц возможно, произведение $\mathbf{D}\cdot\mathbf{D}^{'}$ является квадратной матрицей, она имеет три строки и три столбца, ее элементы

$$(D \cdot D^{'}) = (c_{ik})_{(m,q)} = \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right)_{(m,q)} \quad \text{,где} \quad D^{'} = (a_{ik})_{(m,n)}, \quad D = (b_{i,k})_{(p,q)},$$

причем $a_{i,k}$ – элементы матрицы $\mathbf{D}^{'}$, имеющей три строки (m=3) и два столбца (n = 2); b_{ik} – элементы матрицы **D,** имеющей две строки (p=2) и три столбца (q=3). Произведение матриц $\mathbf{D}\cdot\mathbf{D'}$ – квадратная матрица, так как m=q=3.

Для системы воспроизведения, состоящей из $N^{'}$ громкоговорителей, соответственно имеем $m = q = N^{'}$. По сути дела размер матрицы ($\mathbf{D}\cdot\mathbf{D}^{'}$) определяется числом громкоговорителей $N^{'}$ системы воспроизведения.

Определитель квадратной матрицы ($\mathbf{D}\cdot\mathbf{D}^{'}$)

$$\Delta = \sum_{i} c_{ik} \Delta_{ik},$$

где Δ_{ik} – ее алгебраические дополнения, в данном случае

$$\begin{aligned} \Delta_{11} &= c_{22}c_{33} - c_{23}c_{32}, \\ \Delta_{12} &= (-1)(c_{21}c_{33} - c_{23}c_{31}), \\ \Delta_{13} &= c_{21}c_{32} - c_{22}c_{31}) \quad . \end{aligned} \qquad (1.5)$$

По определению элемент d_{ik} обратной матрицы $(\mathbf{D}\cdot\mathbf{D}^{'})^{-1}$ равен транспонированному алгебраическому дополнению Δ_{ki} соответствующего элемента исходной матрицы ($\mathbf{D}\cdot\mathbf{D}^{'}$), деленному на определитель Δ:

$$d_{ik} = (\Delta_{ki} / \Delta).$$

Транспонированные алгебраические дополнения нетрудно получить из (1.5) перестановкой индексов. При этом матрица $(\mathbf{D}\cdot\mathbf{D}^{'})^{-1}$ не существует, ес-

ли определитель матрицы ($\mathbf{D}\cdot\mathbf{D}'$) равен нулю. И, наконец, вычисляются элементы псевдообратной матрицы $\mathbf{D}^+$ как результат произведения матриц $\mathbf{D}'$ и $(\mathbf{D}\cdot\mathbf{D}')^{-1}$.

Итак, результатом пространственного декодирования $\mathbf{CD} = \mathbf{E}$ при $N' = 2$ или $\mathbf{CD}^+ = \mathbf{E}$ при $N' > 2$ является разделение сигналов $y_j(t)$ друг от друга и выделение информации об уровне сигнала каждого громкоговорителя и о направлении φ_j на него. Начало всей этой совокупности векторов совпадает с точкой расположения слушателя (см. рис. 1.80,а). Направление каждого вектора указывает местоположение в пространстве соответствующего ему громкоговорителя СВ, а его длина зависит от уровня излучаемого сигнала. В качестве примера на рис. 1.80,а показано расположение векторов $\mathbf{\Phi}=\boldsymbol{a_2}$, $\boldsymbol{ЛТ} = \boldsymbol{a_1}$, $\boldsymbol{ПТ} = \boldsymbol{a_3}$ для системы воспроизведения «треугольник». Длины векторов ***Ф, ЛТ, ПТ*** – определяют соотношение уровней сигналов, излучаемых соответствующими громкоговорителями.

Оценка азимута КИЗ формируется на этапе ассоциации формы (второй этап обработки информации в слуховой системе). Для системы воспроизведения, состоящей из N' громкоговорителей, имеем (см. рис. 1.80,*б*):

$$y = \sum_{j=1}^{N'} m_j y_j; \quad m_j = 10^{-0,05 K_j \Delta\tau_j}$$

где y – вектор кажущегося источника звука; y_j – векторный сигнал j-го громкоговорителя; m_j – коэффициент, учитывающий ослабление запаздывающих сигналов в слуховой системе человека; K_j – коэффициент эквивалентности действия на орган слуха значений ΔN_i и $\Delta\tau_i$; $\Delta\tau_j$ – время запаздывания сигнала j-го громкоговорителя относительно опережающего сигнала. Из этого выражения следует, что матрица

$$\mathbf{G} = (m_1, m_2, \ldots, m_j, \ldots, m_{N'})$$

содержит одну строку и

$$\varphi = arctg\left(\sum_{j=1}^{N'} m_j y_j \sin\varphi_j \Big/ \sum_{j=1}^{N'} m_j y_j \cos\varphi_j \right).$$

Здесь φ – азимут кажущегося источника звука; φ_j – азимут j-го громкоговорителя; y_j – амплитуда сигнала j-го громкоговорителя; N' – их общее число. Значения углов отсчитываются относительно медианной плоскости головы слушателя. При этом начало координат в векторной модели совмещено с центром головы слушателя.

Условие сохранения неизменным уровня громкости КИЗ при его перемещении в пространстве выполняется, если

$$\sum_{j=1}^{N'} y_j^2 = const.$$

Теперь назовем ряд общих психофизических закономерностей, подлежащих учету при оценке азимута КИЗ в ассоциативной модели слуха на этапе ассоциации формы:

1) кажущийся источник звука образуется, если сигналы громкоговорителей статистически связаны и величина коэффициента корреляции R между ними превышает некоторое пороговое значение $R_П$;

2) обработка бинауральной пары $Л_б$ и $П_б$ выполняется в полосах, соответствующих критическим полосам слуха; для каждой пары сигналов в каждой такой полосе вычисляется вектор y_j, суждение о направлении локализации КИЗ является результатом сложения полученной совокупности векторов $\{y_j\}_{N'}$;

3) локализация КИЗ является функцией соотношения уровней и временных сдвигов сигналов прямых звуков громкоговорителей, их взаимного расположения в пространстве относительно слушателя. Эти факторы определяют величины и направления векторов y_i. Отделение сигналов прямых звуков от отзвуков оказывается возможным благодаря эффекту предшествования; наличие отражений от поверхностей помещения приводит к появлению гулкости и объемности в звучании, к росту протяженности формируемых КИЗ;

4) выбранное расположение громкоговорителей в пространстве должно обеспечивать разделение векторных сигналов на этапе ассоциации места. Это условие выполняется не всегда. Пусть система воспроизведения содержит два громкоговорителя $Гр_1$ и $Гр_2$ (рис. 1.80,*б*), расположенных зеркально относительно линии базы II – II ушей слушателя. В данном случае значения бинауральных параметров для этой пары громкоговорителей практически равны, приблизительно одинаковы для них и пары коэффициентов пространственного кодирования ($H_{11} \approx H_{12}$ и $H_{21} \approx H_{22}$) матрицы **D**. Поэтому определитель этой матрицы стремится к нулю ($\Delta = H_{11}H_{22} - H_{21}H_{12} \to 0$), а следовательно, обратная матрица $\mathbf{D}^{-1}$ пространственного декодирования в этом случае не существует и разделение сигналов громкоговорителей $Гр_1$ и $Гр_2$ на этапе ассоциации места оказывается невозможным. При $N' > 2$ также возникают ситуации, когда определитель матрицы (**DD'**) равен нулю, тогда псевдообратная матрица $\mathbf{D}^{+}$ не существует и разделить сигналы громкоговорителей СВ на этапе ассоциации места также нельзя. Отсутствие по этой причине полной информации на этапе ассоциации формы делает невозможным в такой СВ образование КИЗ и его плавное перемещение вдоль линий базы громкоговорителей соответствующих пар. Заметим, что данный вывод справедлив для любой пары источников звука,

расположенных зеркально относительно линии II-II (рис. 1.80,*б*). Отсутствие всей полноты информации, необходимой для обработки слуховых сигналов на этапе ассоциации формы, делает невозможной образование КИЗ и его плавное перемещение вдоль линии баз громкоговорителей $Гр_1$ и $Гр_3$ или $Гр_2$ и $Гр_4$ системы воспроизведения типа «квадрат». Здесь возможно лишь скачкообразное перемещение КИЗ из позиции одного громкоговорителя в позицию другого (на линиях боковых баз по отношению к слушателю). При близких же уровнях сигналов этих громкоговорителей возникает ощущение неопределенности в локализации. Именно по этой причине система воспроизведения не должна содержать пар громкоговорителей, расположенных зеркально относительно линии базы (II – II) ушей слушателя. С этих позиций системы воспроизведения типа «квадрат» и «параллелепипед» и расположение слушателя в точке симметрии последних не могут быть признаны удачными, так как локализация КИЗ на линиях боковых баз громкоговорителей окажется невозможной. Этот вывод подтвержден практикой;

5) величины $m_1, m_2, \ldots, m_N'$, образующие вектор-строку учитывают особенности обработки сигналов в слуховой системе человека. В отсутствие корреляционной связи сигнала *j*-того действительного источника звука – громкоговорителя – соответствующий ему коэффициент m_i принимает значение, равное 0, и этот сигнал при образовании КИЗ на этапе ассоциации формы не учитывается. Для сигналов коррелированных источников звука величины m_j не равны 0. Напомним, что временной сдвиг $\Delta\tau$ между коррелированными сигналами, поступающими от громкоговорителей СВ, трансформируется в слуховой системе в соответствующее изменение их уровня. При этом величина ослабления уровня каждого запаздывающего сигнала может быть рассчитана с помощью коэффициента эквивалентности *К*. Этот производимый в слуховой системе «обмен» времени на интенсивность должен приводить к изменению величины вектора $m_j y_j$ сигнала, соответствующего *j*-му громкоговорителю. При этом уровень запаздывающего сигнала в результате этого «обмена» уменьшается, поэтому величина m_j должна быть меньше 1, но больше 0.

В простейшем случае, когда расстояния от слушателя до громкоговорителей СВ одинаковы, а излучаемые ими сигналы отличаются только по уровню ($\Delta\tau$=0), значения параметров m_j в первом приближении могут быть приняты равными 1. Покажем справедливость этого заключения для данного частного случая. Если допустить, что при оценке азимута источника звука (особенно в области нижних частот F < 600 Гц) решающую роль оказывает значение бинауральной временной разности $\Delta\tau_б'$, то для указанного здесь простейшего случая справедливо соотношение

$$\Delta\tau_б^{'} = \frac{d_{экв}}{c_{зв}} = \frac{y_1 Sin\varphi_1 + y_2 Sin\varphi_2 + ... + y_{N'} Sin\varphi_{N'}}{y_1 + y_2 + ... + y_{N'}},$$

где $y_1, y_2, \ldots, y_N'$ амплитуды сигналов, излучаемых громкоговорителями системы воспроизведения; $\varphi_1, \varphi_2, \ldots, \varphi_N'$ – значения углов на громкоговорители СВ, отсчитываемые относительно медианой плоскости головы слушателя; N' – число громкоговорителей системы воспроизведения; $d_{ЭКВ}$ – расстояние между фазовыми центрами раскрыва ушных раковин; $c_{зв}$ – скорость звука. Напомним, что в случае одного действительного источника звука величина $\Delta\tau_б$ бинауральной пары сигналов определяется выражением

$$\Delta\tau_б = d_{экв} \sin\varphi / c.$$

Здесь φ – направление на действительный источник звука, как и ранее, отсчитываемое относительно медианной плоскости головы слушателя. Если величины $\Delta\tau_б$ и $\Delta\tau_б'$ равны, то оценки азимута φ кажущегося источника звука, формируемого сигналами $\{y_j x_j(t)\}$ системы воспроизведения, состоящей из N' равноудаленных от слушателя громкоговорителей, и единственного действительного источника звука должны совпадать. Отсюда следует

$$\sin\varphi = \frac{y_1 Sin\varphi_1 + y_2 Sin\varphi_2 + ... + y_{N'} Sin\varphi_{N'}}{y_1 + y_2 + ... + y_{N'}}, \qquad (1.6)$$

где φ - оценка азимута КИЗ в системе воспроизведения, состоящей из N' равноудаленных от слушателя громкоговорителей. Заметим, что при $N' = 2$ и $\varphi_1 = -\varphi_2$ (случай симметричного расположения громкоговорителей относительно медианной плоскости головы слушателя) имеем

$$\sin\varphi = \sin\varphi_1 (y_1 - y_2)/(y_1 + y_2).$$

Это равенство известно в стереофонии под названием «*закон синусов*».

Далее, если учесть, что при оценке азимута слушатель, во-первых, «поворачивает» голову в направлении φ кажущегося источника звука и, во-вторых, совершает непроизвольно вращательные движения головой около этого направления с амплитудой $\psi \to 0$, то выражение (1.6) может быть преобразовано к виду

$$tg\varphi = \frac{y_1 Sin\varphi_1 + y_2 Sin\varphi_2 + ... + y_{N'} Sin\varphi_{N'}}{y_1 \cos\varphi_1 + y_2 \cos\varphi_2 + ... + y_{N'} \cos\varphi_{N'}}. \qquad (1.7)$$

Выражение (1.7) соответствует представлениям векторной модели локализации КИЗ, если $m_j = 1$. При $N' = 2$, $\varphi_1 = -\varphi_2 = 30^0$ для значений углов $\varphi_1 \leq 30^0$ с достаточной для практики точностью оценка азимута КИЗ может быть рассчитана по формуле

$$\varphi \approx 0,58[(y_1 - y_2)/(y_1 + y_2)],$$

что хорошо согласуется с данными профессора Я.А.Альтмана. Это равенство подтверждается экспериментом: в области малых значений углов $\varphi_1 \leq 30^0$ на громкоговорители из точки расположения слушателя оценка азимута КИЗ при интенсивностной стереофонии и симметричном расположении слушателя относительно последних определяется только соотношением уровней сигналов громкоговорителей и не зависит от расстояния y до линии базы $Гр_1$ и $Гр_2$.

В случае, когда расстояния до громкоговорителей не одинаковы и излучаемые ими сигналы отличаются как по уровню, так и по времени запаздывания для преодоления затруднений в оценке азимута КИЗ необходимо, используя понятие коэффициента эквивалентности, перейти от смешанной стереофонии к чисто интенсивностной и лишь после этого воспользоваться выражением (1.7).

И последнее. Проблема повышения качества звучания матричных систем (рис. 1.78) требует поиска оптимальных структур матриц **A** и **B** и системы воспроизведения, обеспечивающих передачу пространственной информации в максимальном объеме при минимальных величинах пространственных искажений. Работа устройств, определяющих структуру этих матриц, должна рассматриваться во взаимосвязи с учетом условий прослушивания и свойств пространственного слуха человека. При этом имеющаяся в сигналах Л(t) и П(t) информация о пространственном размещении звуковых образов в стереопанораме, их количестве и другие возможные сведения должны быть использованы для управления процессом их декодирования с целью получения наиболее четких и уверенно локализуемых КИЗ, максимально возможного размера области уверенной локализации КИЗ и зоны стереофонического эффекта. Именно комплексный учет всей этой совокупности факторов должен выполняться при разработке эффективных алгоритмов декодирования сигналов Л(t) и П(t) в матричных звуковых системах. Процесс управления декодированием сигналов стереопары не должен быть заметен на слух.

Предельно-достижимым качеством звучания в матричных звуковых системах является то, которое обеспечивается в многоканальной стереофонической системе звукопередачи с числом раздельных каналов $n = N'$ и с идентичной системой воспроизведения. Такая система звукопередачи явля-

ется для матричной системы эталоном. Наилучшими возможностями передачи пространственной информации обладает система воспроизведения типа "трапеция" (рис. 1.81).

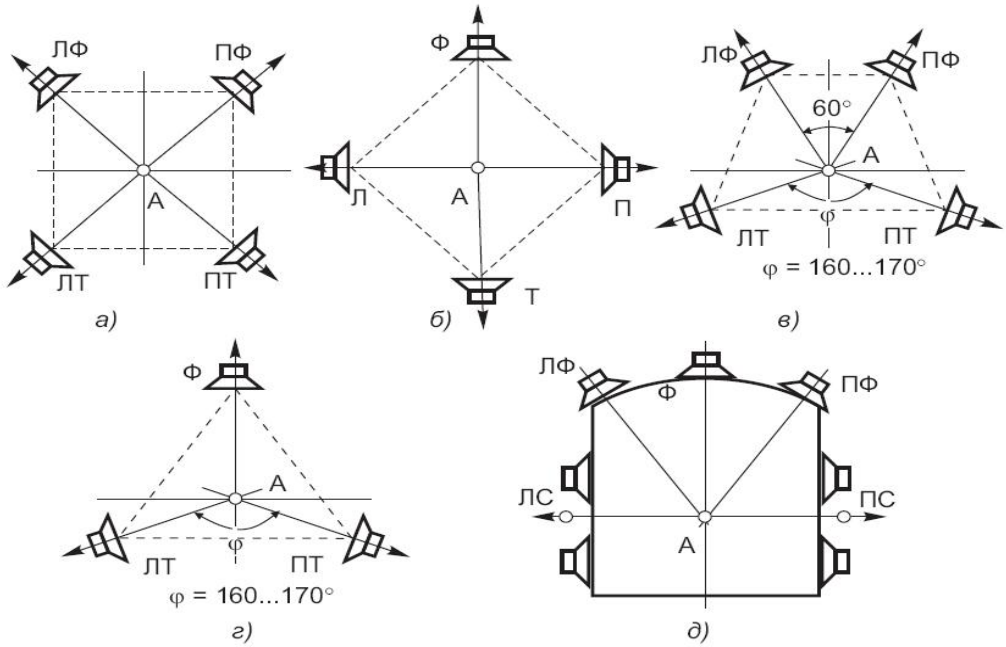


Рис. 1.81. Разновидности систем воспроизведения: *а* – квадрат; *б* – ромб; *в* – трапеция; *г* – треугольник; *д* – принятая в кинематографе; Л, ЛФ, Ф, ПФ, П, ЛТ, ПТ – соответственно левый, левый фронтальный, фронтальный, правый фронтальный, правый, левый тыловой и правый тыловой громкоговорители; ЛС и ПС – громкоговорители левой и правой стен кинозала

1.12. Бинауральная демаскировка источников звука

Важнейшим свойством пространственного слуха человека, в значительной степени определяющим прозрачность звучания, является бинауральная демаскировка.

Напомним, что демаскировкой называют снижение порога маскировки при выделении отдельных сигналов (источников звука) из одновременно действующей на слушателя их совокупности. Это достигается путем соответствующей дополнительной обработки сигналов, как в периферийном, так и в центральном отделах слуховой системы. Заметим, что звуковые сигналы имеют перекрывающиеся спектры, поэтому классическая теория фильтров с ее областями пропускания и затухания здесь оказывается непригодной. Ухо является в этом смысле гораздо более тонким инструментом.

Прозрачность звучания определяется способностью слушателя разделять воспринимаемые сигналы, используя их упорядоченность по форме

и в пространстве. Приведем следующий пример. Представим себе несколько прозрачных контурных рисунков животных, наложенных друг на друга. В этой ситуации разделение и последующее опознавание животных становится возможным только благодаря различию их форм. Термин "форма" имеет здесь тот же самый смысл, что и в теории сигналов, а сама эта ситуация эквивалентна монофонической передаче, при которой все инструменты ансамбля локализуются слушателем в одной точке – позиции громкоговорителя. Следуя взглядом за линией одного из рисунков, мы тем увереннее выбираем путь дальнейшего следования, чем более отчетливо чувствуется (опознается) форма животного, т.е. связь уже пройденного пути с дальнейшим его продолжением.

Очевидно, что разделение совмещенных в пространстве рисунков тем сложнее, чем ближе формы животных (тембры и ритмы звучаний). Если же эти рисунки разнести в пространстве (пусть даже на небольшой угол), то данная задача решается значительно проще. Точно также разнесение источников звука в пространстве является определяющей причиной повышенной прозрачности звучания, особенно в ситуации, когда тембры и ритмы звучаний близки. Слушатель, желая выделить из общего состава звучания партию какого-либо инструмента, концентрирует свое внимание в направлении его расположения в пространстве, что приводит к уменьшению маскирующего действия звуков, воспринимаемых с других направлений. За счет этого повышается отношение сигнал-помеха для выделяемого источника звука и, как следствие этого, улучшается прозрачность звучания. Влияние пространственного разнесения источников звука на изменение условий разделимости соответствующих им сигналов можно оценить по изменению величины артикуляции или порога слышимости выделяемого источника звука.

По мнению многих исследователей, бинауральное освобождение от маскировки предполагает использование слуховой системой временных $\Delta\tau_б$ и интенсивностных $\Delta N_б$ различий бинауральной пары сигналов, соответствующих отдельным источникам звука для разделения их друг от друга.

Покажем это, используя метод артикуляции. Для этого в качестве исходных сигналов выберем монофонические записи трех различных по содержанию речевых отрывков, произнесенных одними тем же человеком в одинаковом ритме и, по возможности, с одинаковой интонацией. Это условие позволяет практически устранить влияние тембра и характера исполнения на результаты эксперимента. Предложим эксперту оценить разборчивость одного из источников звука при мешающем действии двух других. При этом для оценки разборчивости выделяемого источника используем специальный речевой текст. Он содержит 130 слов, записанных при чтении артикуляционных таблиц. Два других (маскирующих) речевых источника – повторяющиеся отрывки дикторского текста длительностью звучания

7…10 с. Заметим, что в начале эксперимента слушатели затрудняются распознавать слова (или слоги) артикуляционной таблицы, их внимание непроизвольно переключается на разные звуковые объекты. Однако через несколько секунд слушателям удается сосредоточить свое внимание на выделяемом источнике звука и распознаваемость речи резко увеличивается. Поэтому оценивалась разборчивость последних 100 слов. Подобная методика в несколько измененном виде впервые предложена в [1.69].

Результаты этих исследований представлены на рис. 1.82 для стереопанорамы, состоящей из трех источников, отдельно для чисто интенсивностной (*а*) и чисто временной (*б*) стереофонии. Здесь по оси ординат отложена

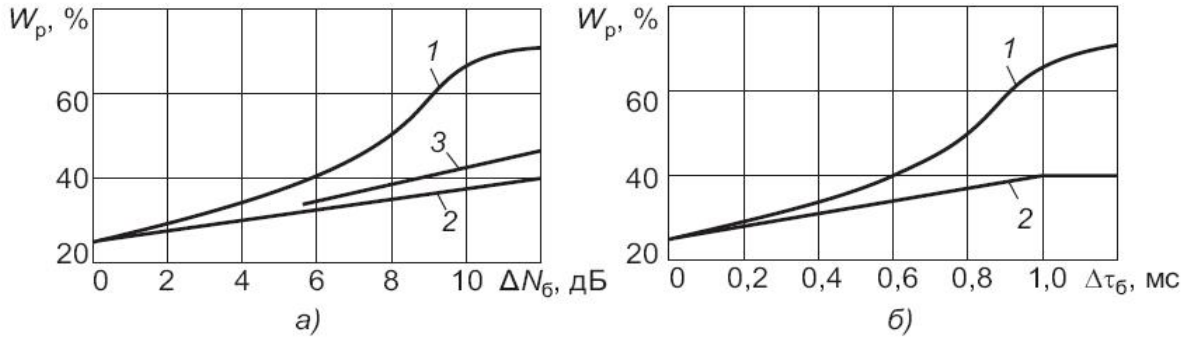


Рис. 1.82.Изменение артикуляции речевых источников пространственной панорамы при интенсивностной (*а*) и временной (*б*) стереофонии: размер базы громкоговорителей 1,8 м; угол между направлениями на громкоговорители 60^0

слоговая разборчивость $W_р$, %, для боковых (кривая 1) и центрального (кривая 2) источников; по оси абсцисс – разность уровней $\Delta N_б$ и временной сдвиг $\Delta\tau_б$ пары сигналов, формирующих боковые КИЗ. Заметим, что при проведении данного эксперимента центральный источник звука для слушателя, расположенного на оси симметрии громкоговорителей, всегда оставался в центре базы (для него значения $\Delta N_ц$ или $\Delta\tau_ц$ равнялись нулю), в то время как боковые источники сдвигались влево и вправо от центрального источника на одинаковое расстояние каждый. Для них вводимые значения $\Delta N_б$ (или $\Delta\tau_б$) были равны по величине, но противоположны по знаку. Данные экспериментальные исследования показали [1.19], что при $\Delta N_б = \Delta N_ц = 0$ (или $\Delta\tau_б = \Delta\tau_ц = 0$) слушатели воспринимают все три КИЗ из одного направления (центра базы громкоговорителей). При этом разборчивости каждого из источников минимальны и одинаковы по величине. При $|\Delta N_б|$=2…4 дБ и $\Delta N_ц = 0$ (или $|\Delta\tau_б| = 0{,}2…0{,}4$ мс и $\Delta\tau_ц = 0$) слушатель воспринимает размытый (объемный) звуковой образ без возможности четкого пространственного разделения отдельных источников звука. Незначительное повышение разборчивости речевых источников свидетельствует о том, что в этом случае условии для разделения формирующих их сигналов несколько улучшаются. Дальнейшее увеличение $|\Delta N_б|$ (или $|\Delta N_б|$) значительно улучшает возможность пространственного разделения боковых (левого и

правого) источников. Словесная разборчивость последних уже при $|\Delta N_б|$=10…12 дБ (или $|\Delta\tau_б|$ = 1…1,2 мс) составляет практически 100%. На каждом из боковых источников слушатели могут легко сосредоточить свое внимание и прослушать соответствующий текст. Несколько иная картина наблюдается для центрального КИЗ: существенного увеличения его разборчивости не наблюдается (кривая 2). Это явление субъективно воспринимается слушателями как «провал середины».

Покажем, что именно пространственное разнесение источников – основная причина улучшения условий дляих разделения. Для этой цели предложим слушателю два варианта звуковой панорамы. В первом случае для всех трех пар сигналов значения ΔN и $\Delta\tau$ выберем равными нулю. Очевидно, что при этом все три КИЗ локализуются слушателем, находящимся на оси симметрии громкоговорителей, в одной точке посередине базы. Во втором случае величины одновременно вводимых для каждой пары сигналов значений ΔN и $\Delta\tau$ выберем так, чтобы они все по-прежнему локализовались бы в центре базы громкоговорителей. Оказывается, что и в этом случае их разборчивость останется одинаковой: расхождение результатов не превышает 10…15%.

Изложенное подтверждает, что именно пространственное разнесение источников звука, формируемых одинаковыми по форме сигналами, улучшает условия для их разделения за счет механизма бинаурального освобождения от маскировки (бинауральной демаскировки).

1.13. Модели бинауральной демаскировки

Накопленные сведения в области физиологии слуха, особенности построения его периферийных и центральных отделов, новейшие знания субъективного поведения экспертов при оценке порогов маскировки в различных экспериментальных условиях и, наконец, известные из классической теории методы обнаружения и выделения сигналов из помех позволили разработать ряд моделей бинаурального освобождения от маскировки.

Наиболее известны из них три: модель накопления [1.53, 1.54], корреляционная *ЕС*-модель [1.55, 1.56] и модель корреляционного пеленгования [1.19].

Модель накопления

Структурная схема модели представлена на рис. 1.83. Она содержит: набор полосовых фильтров $ПФ_υ$, разделяющих входные сигналы $y_1(t)$ и $y_2(t)$ на полосы, соответствующие по ширине критическим полосам слуха; суммирующие (+) и вычитающие (-) устройства; квадратирующие устрой-

ства (15-20); интеграторы (21-26) и устройства (27-32), вычисляющие на каждом из выходов отношение мощности полезного сигнала $P_{s\upsilon}$ к полной мощности маскирующих сигналов $P_{n\upsilon}$. Здесь же (для большей наглядности) частично показаны операции, выполняемые над сигналами при прохождении последними каждого из блоков модели.

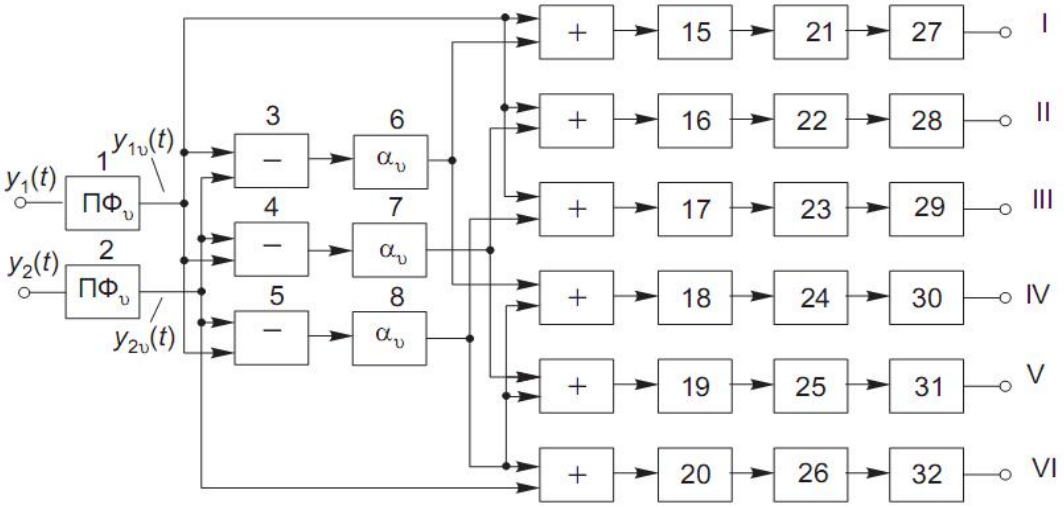


Рис. 1.83. Структурная схема модели накопления

Кратко рассмотрим принцип ее работы. Пусть на каждый из входов модели (рис. 1.83) воздействует сумма только двух сигналов: полезного и мешающего, так что

$$y_1(t) = a_{1S} f_S(t - \tau_{1S}) + a_{1n} f_n(t - \tau_{1n})$$
$$y_2(t) = a_{2S} f_S(t - \tau_{2S}) + a_{2n} f_n(t - \tau_{2n}),$$

где индексы 1 и 2 соответствуют левому и правому входам модели (органа слуха); индекс s – полезному сигналу; индекс n – мешающему (маскирующему) сигналу; a_{1S}, a_{2S}, a_{1n}, a_{2n} – амплитуды полезного и мешающего сигналов для левого и правого уха слушателя; τ_{1S}, τ_{2S}, τ_{1n}, τ_{2n} – времена запаздывания этих сигналов; $f_S(t)$ и $f_n(t)$ – функции времени, отображающие полезный и мешающий сигналы. Будем считать, что

$$-1 \le f_S, f_n \le 1.$$

В данной модели учтена одна из замечательных особенностей, присущая слуховому анализатору человека, – разделять (уже в периферическом отделе) всю совокупность входных сигналов по частоте на области, называемые критическими полосами (частотными группами) слуха. Эта процедура выполняется гребенкой полосовых фильтров ПФ$_\upsilon$, имеющих полосы прозрач-

ности, равные критическим полосам слуха. Поэтому на выходах этих фильтров имеем

$$y_{1\upsilon}(t) = a_{1S} f_{S\upsilon}(t - \tau_{1S}) + a_{1n} f_{n\upsilon}(t - \tau_{1n})$$
$$y_{2\upsilon}(t) = a_{2S} f_{S\upsilon}(t - \tau_{2S}) + a_{2n} f_{n\upsilon}(t - \tau_{2n}).$$

Здесь индекс υ определяет номер частотной группы слуха. Последующая обработка сигналов выполняется отдельно внутри каждой критической полосы слуха. В начале эти сигналы подвергаются суммарно-разносным преобразованиям. При этом накопление полезного сигнала, вследствие выполнения этой процедуры, происходит на одном из выходов сумматоров, но каком именно заранее не известно. Поэтому дальнейшей обработке подвергается каждый из шести полученных таким образом выходных сигналов (блоки 9-14). При этом учитываются следующие сведения, объясняющие последующее формирование порогов слышимости:

-абсолютный порог слышимости при восприятии образуется слухом путем оценки не амплитуды, а мощности (энергии), сосредоточенной внутри каждой частотной группы; если в какой-либо из них мощность (энергия) сигнала превысит пороговое значение, то сигнал будет услышан;

-при определении относительного порога слышимости чистого тона, маскируемого шумом, можно пренебречь всеми компонентами последнего, лежащими за пределами образованной слухом критической полосы. Тон также будет услышан, когда его интенсивность достигнет определенного порогового значения по отношению к интенсивности шума, приходящейся на эту частотную группу.

Отсюда следует, что для каждого из упомянутых выше шести сигналов должна быть вычислена мощность, что требует их возведения в квадрат (блоки 15-20) после выполнения суммарно-разностных преобразований и последующего их усреднения во времени (блоки 21-26). По мнению ряда исследователей, интервал усреднения должен составлять ≈ 200 мс. После усреднения должны быть отброшены все члены, где под знаком интеграла стоит произведение некоррелированных функций, которыми являются соответственно полезный (выделяемый) $f_{S\upsilon}(t)$ и мешающий (маскируемый) $f_{n\upsilon}(t)$ сигналы, так как при интервале усреднения равном 200 мс величина интеграла становится очень малой.

Блоки 27-32 (рис. 1.83) рассчитывают для каждого из выходных каналов отношения накопленной мощности полезного сигнала $P'_{S\upsilon}$ к мощности маскирующего сигнала $P'_{n\upsilon}$.

Опуская все промежуточные выкладки, приведем окончательные выражения для расчета этих отношений

$$\left(\frac{P'_{S\upsilon}}{P'_{n\upsilon}}\right)_I = \frac{[4a_{1S}^2 + a_{2S}^2 + 4a_{1S}a_{2S}R_\upsilon(\Delta\tau_S)]P_{S\upsilon}}{[4a_{1n}^2 + a_{2n}^2 + 4a_{1n}a_{2n}R_\upsilon(\Delta\tau_n)]P_{n\upsilon}}; \quad \left(\frac{P'_{S\upsilon}}{P'_{n\upsilon}}\right)_{II} = \frac{[4a_{1S}^2 + a_{2S}^2 - 4a_{1S}a_{2S}R_\upsilon(\Delta\tau_S)]P_{S\upsilon}}{[4a_{1n}^2 + a_{2n}^2 - 4a_{1n}a_{2n}R_\upsilon(\Delta\tau_n)]P_{n\upsilon}}$$

$$(\frac{P'_{S\upsilon}}{P'_{n\upsilon}})_{III} = \frac{a_{2S}^2 P_{S\upsilon}}{a_{2n}^2 P_{n\upsilon}} \; ; \; (\frac{P'_{S\upsilon}}{P'_{n\upsilon}})_{IV} = \frac{[a_{1S}^2 + 4a_{2S}^2 + 4a_{1S}a_{2S}R_\upsilon(\Delta\tau_S)]P_{S\upsilon}}{[a_{1n}^2 + 4a_{2n}^2 + 4a_{1n}a_{2n}R_\upsilon(\Delta\tau_n)]P_{n\upsilon}} \; ; \; (\frac{P'_{S\upsilon}}{P'_{n\upsilon}})_{V} = \frac{a_{1S}^2 P_{S\upsilon}}{a_{1n}^2 P_{n\upsilon}}$$

$$(\frac{P'_{S\upsilon}}{P'_{n\upsilon}})_{VI} = \frac{[4a_{2S}^2 + a_{1S}^2 - 4a_{1S}a_{2S}R_\upsilon(\Delta\tau_S)]P_{S\upsilon}}{[4a_{2n}^2 + a_{1n}^2 - 4a_{1n}a_{2n}R_\upsilon(\Delta\tau_n)]P_{n\upsilon}} .$$

Здесь цифры I, II, III, IV, V, VI определяют номер выхода модели (рис. 1.83); $\Delta\tau_S=(\tau_{1S} - \tau_{2S})$ и $\Delta\tau_n= (\tau_{1n} - \tau_{2n})$ – бинауральные временные разности соответственно для пар полезного и мешающего сигналов; $R_\upsilon(\Delta\tau_S)$ и $R_\upsilon(\Delta\tau_n)$ – значения нормированной функции корреляции соответственно для полезного и маскирующего сигналов; $P_{S\upsilon}$ и $P_{n\upsilon}$ – средние за время T (интервал усреднения) значения мощностей для нормализованных функций полезного $f_{S\upsilon}(t)$ и маскирующего $f_{n\upsilon}(t)$ сигналов критической полосе слуха υ.

Если в качестве полезного и маскирующего сигналов используется белый шум, а верхняя и нижняя частотные границы для критической полосы слуха υ составляют F_1 и F_2, то можно написать, что

$$R_\upsilon(\Delta\tau_S) = \frac{\sin(2\pi F_2\Delta\tau_S) - \sin(2\pi F_1\Delta\tau_S)}{2\pi\Delta F_\upsilon\Delta\tau_S} \; ; \; R_\upsilon(\Delta\tau_n) = \frac{\sin(2\pi F_2\Delta\tau_n) - \sin(2\pi F_1\Delta\tau_n)}{2\pi\Delta F_\upsilon\Delta\tau_n} ,$$

где ΔF_υ – ширина критической полосы υ.

Как только величина отношения $P'_{S\upsilon} / P'_{n\upsilon}$ на каком либо из шести выходов в пределах критической полосы слуха превысит пороговое значение, то относительный порог слышимости будет достигнут и полезный сигнал будет услышан при его восприятии на фоне мешающих звуков. И последнее. При формировании окончательного суждения необходимо также учесть взаимное влияние эффектов маскировки в соседних частотных группах друг на друга. Поэтому в модели накопления результату, полученному отдельно для каждой критической полосы, приписывается определенный весовой множитель α_υ, изменяющийся по величине при переходе от одной критической полосе слуха к другой (рис. 1.84, где υ- номер частотной группы).

В качестве примера на рис. 1.85 показано снижение порога маскировки

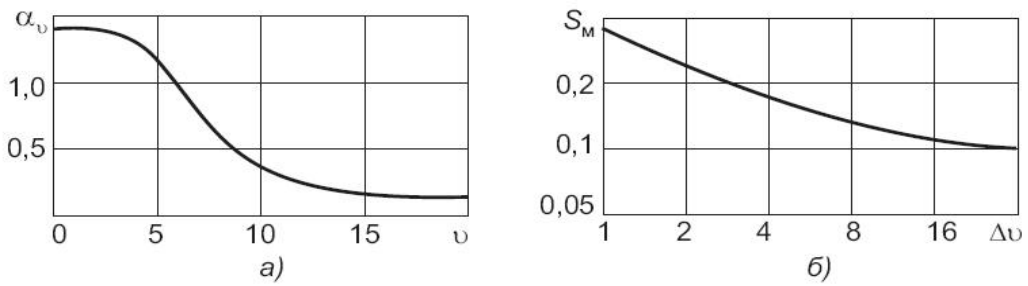


Рис. 1.84. Примеры весовых функций, учитывающих: *а* - взаимное влияние частотных групп друг на друга; *б* – изменение пороговой величины $S_М$ от числа частотных групп с одинаковым отношением энергии полезного сигнала к шуму

тона частотой 200 Гц от разности фаз его бинауральной пары сигналов (что эквивалентно введению $\Delta\tau_S$) маскируемого белым шумом с уровнем 70 дБ, для которого ΔN_n и $\Delta\tau_n$ равны нулю. По оси ординат здесь отложена разность уровней тона $N_{0/0}$ и тона $N_{\varphi/0}$ на пороге слышимости, то есть $\Delta N = N_{0/0} - N_{\varphi/0}$, дБ. Видно, что результаты вычислений (сплошная кривая) хорошо подтверждаются данными эксперимента (пунктир на рис. 1.85).

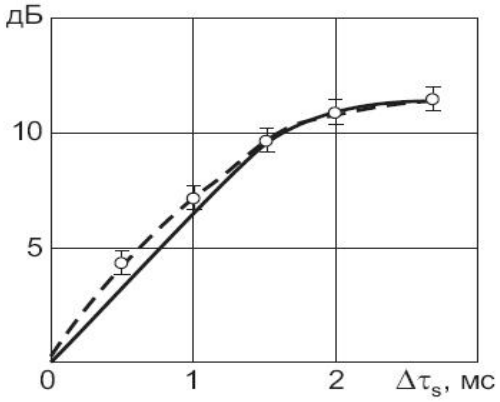


Рис. 1.85.Снижение порога маскировки чистого тона частотой 200 Гц от величины его бинауральной временной разности

Корреляционная *ЕС*-модель

Согласно корреляционной модели пространственного слуха [1.58;1.59] при бинауральном воздействии пары коррелированных сигналов, отличающихся по уровню или по времени поступления на левый и правый входы органа слуха, происходит возбуждение трех концептуальных поверхностей, участвующих в процессе слияния звуков. Причем две из них представляют собой плоскости (или поля) текущей автокорреляции r_{a1} и r_{a2} воспринимаемых слушателем сигналов. На центральной концептуальной поверхности реализуется текущая взаимная корреляция $r_{вз}$ входных сигналов. Введение текущей оценки необходимо потому, что реальные сигналы представляют собой случайные функции.

Возбуждения трех концептуальных поверхностей сливаются в единый очаг возбуждения, пространственное положение которого в слуховом центре головного мозга слушателя связано с направлением на источник звука. Анализ звука, разделение сигналов и восприятие направления чрезвычайно тесно связаны между собой. Это отмечается многими исследователями. В силу этого модель бинауральной демаскировки должна содержать устройства, вычисляющие функции автокорреляции и взаимной корреляции входных сигналов. Естественно также предположить [1.58,1.59], что слух при выделении источника звука на фоне мешающих сигналов реагирует («настраивается») на оптимальную задержку $\Delta\tau_{опт} = -\Delta\tau_S$, при которой наступает максимум функции взаимной корреляции для полезного сигнала. При этом данная процедура может выполняться с определенной погрешностью δ. Структурная схема этой корреляционной части модели представлена на рис. 1.86,*а*. Здесь y_1(t) и y_2(t) – бинауральная пара сигналов; $ПФ_\upsilon$ – набор полосовых фильтров, с помощью которых входные сигналы разделяются на

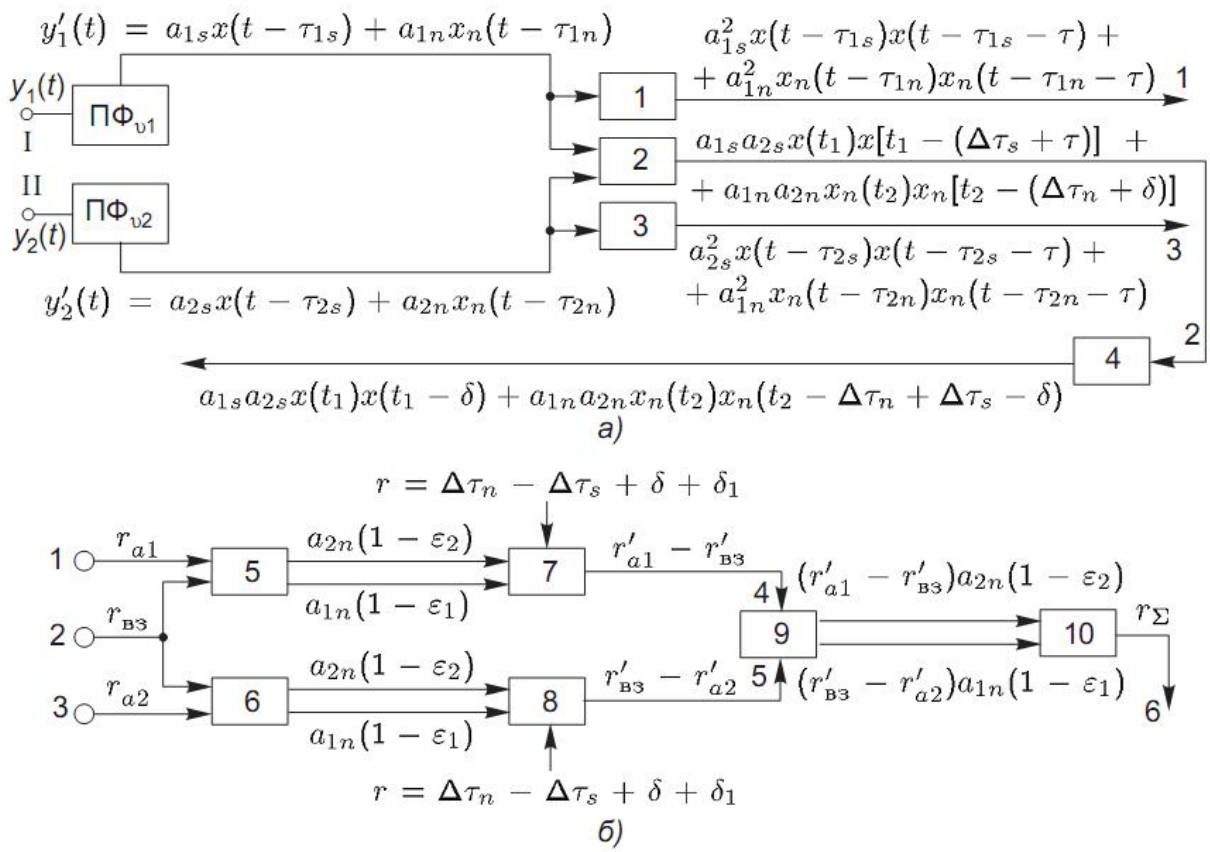


Рис. 1.86.Структурная схема корреляционной *ЕС*-модели бинауральной демаскировки сигналов: а – первая (корреляционная) часть; б – вторая (*ЕС*) часть

сигналы $y_{1\upsilon}$(t) и $y_{2\upsilon}$(t), полосы частот которых соответствующие частотным группам слуха; 1,2, 3 – устройства, вычисляющие автокорреляционные (r_{a1} и r_{a2}) и взаимно корреляционную функции $r_{вз}$ сигналов $y_{1\upsilon}$(t) и $y_{2\upsilon}$(t); 4 –

элемент, вводящий временной сдвиг $\Delta\tau_{опт} = -\Delta\tau_S$ оптимальный по величине для полезного сигнала, но выполняемый с определенной погрешностью δ.

Полученные на ее выходах 1,2,3 сигналы для еще большего улучшения условий разделимости обрабатываются дополнительно во второй части модели (рис. 1.86,*б*), в основу работы которой положен принцип «уравнивания и уничтожения» (*equalization and cancelation – EC*) [1.55,1.56] компонент маскирующих сигналов.

Эта часть модели содержит корректоры сигналов 5,6,9 и вычитающие устройства 7, 8 и 10. Суть *E*-преобразования (выполняемого с помощью корректоров сигналов) состоит в том, чтобы путем введения (при попарной обработке) определенных величин временных и интенсивностных разностей, одинаковых для выделяемого и маскирующего сигналов, уравнять, по возможности компоненты мешающих сигналов, поступающих на входы 1,2; 2,3 и 4, 5 (рис. 1.86,*б*). Если после этой коррекции произвести вычитание соответствующих пар сигналов (*C*-преобразование), то получим в идеале полное уничтожение мешающих сигналов и, следовательно, увеличение отношения сигнал-помеха. Однако предполагается, что *E*-преобразования не могут быть выполнены идеально: при этом всегда возникают хаотически действующие погрешности. Поэтому на выходе каждого вычитающего устройства 7,8,10 будет существовать маскирующий сигнал и отношение сигнал-помеха будет иметь конечную величину. На рис. 1.86 приведены требуемые коэффициенты коррекции сигналов, подвергающихся Е-преобразованию, где ε_1 и ε_2 –погрешности коррекции по интенсивности, а δ и δ_1 – погрешности коррекции по времени.

Теперь, если проделать все операции, предусмотренные второй частью корреляционной *EC*-модели, то можно получить выражение для расчета мощностей полезной и мешающей компонент бинаурального сигнала. При этом, выполняя усреднение погрешностей, возникающих в модели при обработке сигналов, будем считать, что случайные величины ε_1 , ε_2, δ и δ_1 центрированные, статистически независимые и имеют нормальное распределение. Кроме того, для упрощения математических преобразований, предположим, что:

$\sigma_{\varepsilon 1}^2 = \sigma_{\varepsilon 2}^2 = \sigma_\varepsilon^2$ и $\sigma_{\delta 1}^2 = \sigma_\delta^2$, где σ_δ^2 и σ_ε^2 - дисперсии упомянутых выше нормальных процессов;

временные функции $f_{Sv}(t)$ и $f_{nv}(t)$ сигналов представляют собой белый шум.

Проделав все достаточно сложные математические преобразования, связанные с усреднением погрешностей ε_1 , ε_2, δ , δ_1, δ_2 , а также учитывая, что сами величины этих погрешностей достаточно малы, можно получить выражения для расчета мощности нормализованного сигнала и мешающего шума в критической полосе слуха для случая воздействия двух сигналов, один из которых является полезным, а второй маскирующим

$$\langle P_{S\upsilon}\rangle \cong P_\upsilon[2a_{1S}a_{2S}a_{1n}a_{2n} - (1+\sigma_\varepsilon^2)(a_{2S}^2a_{1n}^2 + a_{1S}^2a_{2n}^2)R_\upsilon(\Delta\tau_n)];\ \langle P_{n\upsilon}\rangle \cong 2P_\upsilon a_{1n}^2a_{2n}^2R_\upsilon(\Delta\tau_n)\sigma_\varepsilon^2,$$

где P_υ – мощность нормированного шума в критической полосе слуха; $\langle P_{S\upsilon}\rangle$ - мощность выделяемого сигнала в критической полосе слуха; $\langle P_{n\upsilon}\rangle$ - мощность мешающего (маскирующего) сигнала в критической полосе слуха; $R_\upsilon(\Delta\tau_n)$- значение функции корреляции нормированного шума; знак $\langle\,\rangle$ означает операцию усреднения, выполненную для случайных величин ε_1 , ε_2 , δ , δ_1, δ_2; σ_ε^2- дисперсия погрешностей ε_1 , ε_2.

Теперь, если на слуховой анализатор воздействует не один, а m некоррелированных мешающих сигналов, то полагаем, что операции коррекции и вычитания выполняются отдельно для каждой пары полезный сигнал - мешающий сигнал, после чего результаты суммируются без учета знака. Далее, если полосы частот ΔF полезного и мешающего сигналов больше критической ΔF_υ, то описанный выше способ обработки сигналов производится отдельно для каждой частотной группы, после чего результаты суммируются.

С учетом изложенного полученные выше выражения преобразуются к виду

$$\langle P_S\rangle \cong P\sum_{\upsilon=1}^{\kappa}\sum_{n=1}^{m}\frac{\Delta F_\upsilon}{\Delta F}\frac{1}{m}\left|2a_{1S}a_{2S}a_{1n}a_{2n} - (1+\sigma_\varepsilon^2)(a_{2S}^2a_{1n}^2 + a_{1S}^2a_{2n}^2)R_\upsilon(\Delta\tau_n)\right|;$$

$$\langle P_n\rangle \cong 2\sum_{\upsilon=1}^{k}\sum_{n=1}^{m}\frac{\Delta F_\upsilon}{\Delta F}\left|a_{1n}^2a_{2n}^2R_\upsilon(\Delta\tau_n)\sigma_\varepsilon^2\right|,$$

где P – полная мощность нормированного сигнала в полосе $\Delta F = F_2$-F_1; ΔF_υ – ширина критической полосы слуха; k – число частотных групп, образуемых слухом при восприятии данных сигналов; m - число мешающих сигналов; F_1 и F_2 – нижняя и верхняя частоты сигналов. Здесь принято, что все воздействующие сигналы (как полезный, так и мешающие) имеют одинаковые спектры (в данном случае белый шум), что создает наихудшие условия для их разделения.

Очевидно, что отношение сигнал – помеха на выходе 6 корреляционной *ЕС*-модели может быть представлено выражением

$$K = \langle P_S\rangle / \langle P_n\rangle. \qquad (1.8)$$

Чтобы оценить степень бинауральной демаскировки нужно найти выражение

$$Q_j = K / K_j \text{, причем } K_j = a_{1S}^2 / \sum_{n=1}^{m} a_{jn}^2 \qquad (1.9)$$

где K_j – отношение сигнал–помеха для выделяемого сигнала на левом (j=1) и правом (j = 2) входах слуха.

Корреляционная модель предполагает также наличие механизма суждения, который из трех вариантов выходных сигналов (1, 3 и 6) для дальнейшего анализа выбирает тот, который имеет наилучшее отношение сигнал–помеха для полезного сигнала. Очевидно, что если после обработки мы получим $Q_j < 1$, то это значит, что в процессе бинауральной обработки отношение сигнал-помеха ухудшается, демаскировка в этой ситуации оказывается невозможной.

В качестве примера рассмотрим демаскировку источников звука стереопанорамы. Пусть она состоит их трех КИЗ. При этом отдельно рассмотрим два случая:

а) пространственная звуковая панорама создается введением в канальные сигналы, формирующие каждый из трех КИЗ, только временных разностей $\Delta\tau$ и

б) введением только интенсивностных разностей ΔN.

Причем сам метод синтеза стереопанорамы аналогичен случаю, описанному выше в начале этого раздела, где речь шла об оценке прозрачности звучания методом артикуляции. В отличие от ранее описанного будем считать, что исходными сигналами являются три разных шумовых сигнала одинаковых по полосе частот (F_1=80 Гц; F_2=1080 Гц), но полученных от разных генераторов. Это позволяет считать данные сигналы некоррелированными. Воспроизведение сигналов осуществляется с помощью головных телефонов.

Учитывая метод синтеза стереопанорамы выражения (1.8, 1.9) можно преобразовать к виду:

-для интенсивностной стереофонии ($\Delta N \neq 0$ и $\Delta\tau = 0$)

$$K_{\Delta N} = \frac{\sum_{\upsilon=1}^{\kappa}\sum_{n=1}^{m}\frac{\Delta F_\upsilon}{\Delta F}\frac{1}{m}\left|2a_{1S}a_{2S}a_{1n}a_{2n} - (1+\sigma_\varepsilon^2)(a_{2S}^2a_{1n}^2 + a_{1S}^2a_{2n}^2)\right|}{2\sum_{\upsilon=1}^{k}\sum_{n=1}^{m}\frac{\Delta F_\upsilon}{\Delta F}\left|a_{1n}^2a_{2n}^2\sigma_\varepsilon^2\right|} \qquad (1.10)$$

-для временной стереофонии ($\Delta N = 0$ и $\Delta\tau \neq 0$)

$$K_{\Delta\tau} = \frac{\sum_{\upsilon=1}^{\kappa}\sum_{n=1}^{2}\frac{\Delta F_{\upsilon}}{\Delta F}\frac{1}{m}\left|1-(1+\sigma_{\varepsilon}^{2})R_{\upsilon}(\Delta\tau_{n})\right|}{2\sum_{\upsilon=1}^{k}\sum_{n=1}^{2}\frac{\Delta F_{\upsilon}}{\Delta F}\sigma_{\varepsilon}^{2}\left|R_{\upsilon}(\Delta\tau_{n})\right|}$$

причем относительное изменение этих величин с введением ΔN или $\Delta\tau$ можно представить как

$$K'_{\Delta N,\Delta\tau} = \frac{K_{\Delta N\neq 0,\Delta\tau\neq 0}}{K_{\Delta N=0,\Delta\tau=0}}. \qquad (1.11)$$

Зависимости, показывающие изменение величины $K'_{\Delta N}$ в случае чисто интенсивностной стереофонии, представлены на рис. 1.87 и 1.88. По оси ординат отложено относительное изменение отношения сигнал–помеха в дБ, вычисленное по выражениям (1.10 и 1.11) отдельно для боковых (рис. 1.87) и центрального (рис. 1.88) кажущихся источников пространственной панорамы. По оси абсцисс – разность уровней ($\Delta N_{б}$, дБ) сигналов, формирующих боковые источники. Параметром этих кривых является величина дисперсии σ_{ε}^{2}, характеризующая погрешности при обработке сигналов в слуховой системе.

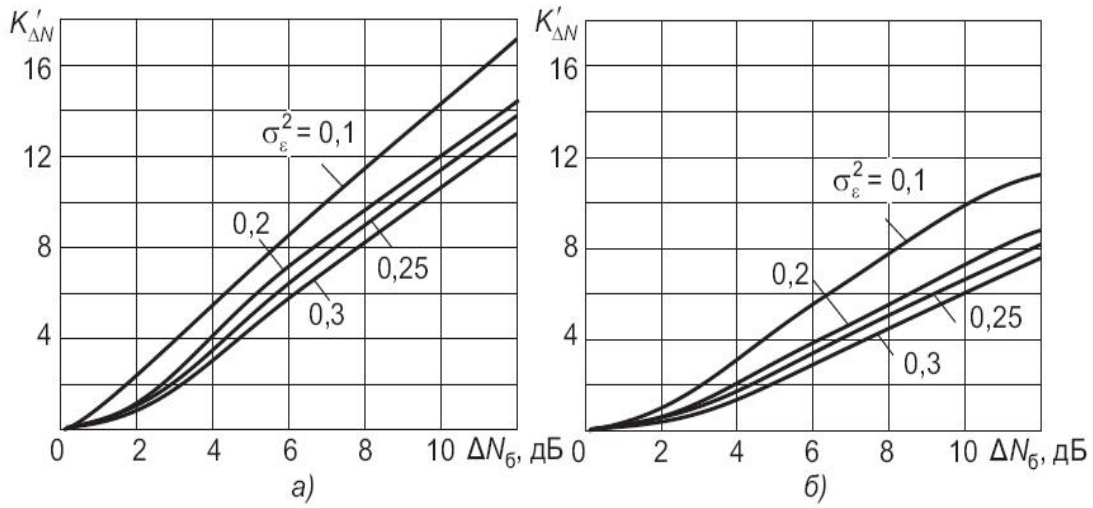


Рис. 1.87. Зависимости отношения сигнал/помеха для боковых (*а*) и центрального (*б*) КИЗ пространственной панорамы при чисто интенсивностной стереофонии: кривая 1 – при σ_{ε}^{2} = 0,1; кривая 2 – при σ_{ε}^{2} = 0,2; кривая 3 – при σ_{ε}^{2} = 0,25; кривая 4 – σ_{ε}^{2} = 0,3

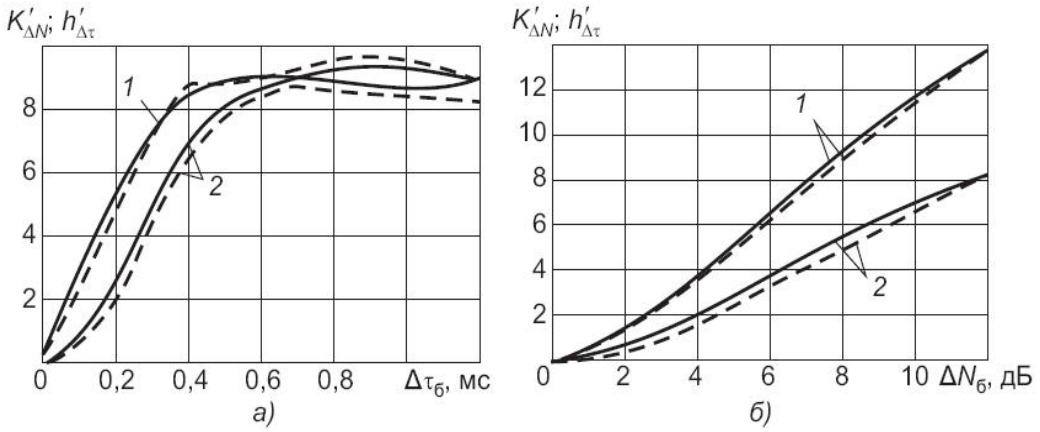


Рис. 1.88. Экспериментальные (пунктирные кривые) и теоретические (сплошные кривые) зависимости (при σ_ε^2 = 0,25) изменения отношения сигнал/помеха для боковых 1 и центрального 2 КИЗ при временной (*а*) и интенсивностной (*б*) стереофонии

Значения σ_δ^2 и $\sigma_{\delta 2}^2$, как и ранее, считаются малыми и по этой причине при расчетах не учитываются. Нетрудно видеть, что отношение сигнал/помеха ($K'_{\Delta N}$) растет тем значительнее, чем больше модуль величины $\Delta N_б$, формирующей боковые кажущиеся источники звука. Кроме того, при увеличении $\Delta N_б$ наблюдается более быстрое изменение отношения $K'_{\Delta N}$ для пар сигналов, формирующих боковые КИЗ. Отсюда следует, что с увеличением пространственного разнесения источников звука должно наблюдаться преимущественное увеличение условий для выделения боковых звуковых образов.

Влияние временной разности $\Delta\tau_б$ на изменение величины $K'_{\Delta\tau}$ для боковых (кривая 1) и центрального (кривая 2) КИЗ представлено на рис. 1.88,*б*, сплошные линии. По оси ординат отложено относительное изменение отношения сигнал – помеха $K'_{\Delta\tau}$, в дБ, вычисленное по выражениям (1.11), по оси абсцисс – временной сдвиг ($\Delta\tau_б$, мс) сигналов, формирующих боковые источники. Эти зависимости (сплошные линии) получены при σ_ε^2 = 0,25. Их характер аналогичен ранее полученным кривым (рис. 1.87, 1.88) только при $\Delta\tau_б$ < 0,6 мс. Дальнейшее увеличение временной разности $\Delta\tau_б$ для сигналов, формирующих боковые КИЗ, не вызывает изменения величин $K'_{\Delta\tau}$, оказывающихся в этой области практически одинаковыми. Это объясняется тем, что при $\Delta\tau_б$ > 0,6 мс, интенсивность, форма и пространственное распределение очагов возбуждений в слуховом центре головного мозга слушателя, соответствующее этим КИЗ, не изменяется. При этом не должны изменяться и условия их разделения соответствующих им сигналов.

Очевидно, что совпадение теоретических и экспериментальных зависимостей, характеризующих условия восприятия центрального и боковых

КИЗ при введении $\Delta N_б$ или $\Delta\tau_б$, должно наблюдаться только при определенном значении σ_ε^2. Найдем это значение, используя метод оценки порога слышимости h. Под величиной h здесь понимается наименьший воспринимаемый уровень выделяемого источника звука, при котором слушатель еще замечает его звучание на фоне мешающих звуков.

Прежде всего, установим количественную связь между величинами $K'_{\Delta N}$, $K'_{\Delta\tau}$ с одной стороны и значениями h с другой. Не вызывает сомнения тот факт, что порогу слышимости будет соответствовать одно и тоже значение величины $K_{\Delta N}$ и $K_{\Delta\tau}$ равное $К_п$. С учетом этого можно написать, что

$$K_П = h^2_{\Delta N} K_{\Delta N} = h^2_{\Delta\tau} K_{\Delta\tau},$$

где $h < 1$, причем $h = (a'_{1S} / a_{1S}) = (a'_{2S} / a_{2S})$. Здесь a'_{1S} и a'_{2S} - амплитуды полезного сигнала, соответствующие порогу слышимости выделяемого источника звука. При ΔN=0 и $\Delta\tau$=0 выражение примет вид

$$K_П = h^2_{\Delta N=0} K_{\Delta N=0} = h^2_{\Delta\tau=0} K_{\Delta\tau=0}. \qquad (1.11,a)$$

После деления (1.11) на (1.11,*а*) и логарифмирования полученные выражения можно записать в виде

$$10\lg\frac{K_{\Delta N\neq 0}}{K_{\Delta N=0}} = 20\lg\frac{h_{\Delta N=0}}{h_{\Delta N\neq 0}}\ ; 10\lg\frac{K_{\Delta\tau\neq 0}}{K_{\Delta\tau=0}} = 20\lg\frac{h_{\Delta\tau=0}}{h_{\Delta\tau\neq 0}}. \qquad (1.12,а)$$

Из (1.12,а) следует, что относительное изменение отношения сигнал/помеха для сигналов выделяемого звукового образа равно по величине относительному изменению его порога слышимости. Следовательно, изменение порога слышимости непосредственно характеризует способность органа слуха человека разделять воспринимаемые сигналы. Очевидно, что эти пороговые величины могут быть найдены экспериментальным путем с помощью методики, изложенной в [1.19]. Эти кривые представлены пунктиром на рис. 1.87 и 1.88. Нетрудно видеть, что при условии, когда σ_ε = 0,5 и $\sigma_\delta \leq 0{,}1$ мс расхождение теории и эксперимента не превышает 10%.

Таким образом, при выбранных условиях корреляционная *ЕС*-модель дает результаты, хорошо согласующиеся с данными опыта и объясняет работу механизма бинауральной демаскировки.

Модель корреляционного пеленгования

В ее основе лежит следующая гипотеза [1.19; 1.58;1.59]:

-положение в пространстве, протяженность и воспринимаемая громкость источника звука определяют место, форму и интенсивность очага возбуждения в слуховом центре головного мозга слушателя;

-при одновременном восприятии нескольких пространственно разнесенных звуковых образов в слуховом центре головного мозга слушателя возникает соответствующее пространственное распределение очагов возбуждений (по-видимому, эта картина является уменьшенной копией исходной пространственной панорамы).

При наложении очагов возбуждений, соответствующих отдельным источникам звука, условия для их разделения, вследствие ограниченной разрешающей способности слуха, оказываются наихудшими, особенно в ситуации, когда звучания близки по тембру и ритмическому рисунку. Бинауральное освобождение от маскировки в этой ситуации оказывается невозможным.

При одновременном восприятии нескольких пространственно разнесенных источников звука в слуховом центре возникает пространственная картина распределения возбуждений, являющаяся отражением реально существующей панорамы. Последнее сопровождается не только улучшением возможностей для анализа сигналов, так как в этом случае используется значительно больший объем слуховой области, но также приводит к снижению взаимной маскировки источников звука. Именно в этом случае работает механизм бинаурального освобождения от маскировки, что приводит к улучшению разделимости соответствующих этим источникам сигналов.

Вся трудность получения критерия, оценивающего изменение условий для разделения источников звука при данном подходе, состоит в нахождении функциональной зависимости, связывающей угловое положение, протяженность и воспринимаемую громкость источника звука, во-первых, с характеристиками формирующего его сигнала и, во-вторых, с пространственным положением, формой и интенсивностью соответствующего этому звуковому образу очага возбуждения.

Весьма подходящей функцией с этой точки зрения является так называемая обостренная функция локализации $r'лок$, полученная путем попарного вычитания составляющих функции локализации (1.1)

$$\begin{aligned} r'лок = &[r_2(\Delta\tau_{22,11} - \Delta\tau) + r_3(\Delta\tau_{21,12} + \Delta\tau)] - \\ &- r_1(\Delta\tau_{12,11}) - r_4(\Delta\tau_{21,22}). \end{aligned} \tag{1.12}$$

Здесь первое слагаемое характеризует возбуждение центральной концептуальной поверхности корреляционной модели слуха, два последних слагаемых – возбуждения соответственно левой и правой концептуальных поверхностей. Возбуждения этих трех поверхностей сливаются в единый

очаг, пространственное положение которого в слуховом центре головного мозга слушателя однозначно связано с направлением на КИЗ. Сама функция локализации (1.1) представляет собой функцию взаимной корреляции бинауральной пары сигналов. Для ее получения, как было показано ранее, можно использовать муляж искусственной головы, в слуховые проходы которой в месте расположения мембраны установлены микрофоны (рис. 1.75).

В модели корреляционного пеленгования в качестве критерия, характеризующего способность слушателя выделять полезный сигнал на фоне маскирующих звуков, может служить, отношение

$$K = (r'_{m\,лок\,\max} / \sum_{i=1}^{n} r'_{i\,лок\,\varphi_m}), \qquad (1.13)$$

где $i = 1,2,..., n$ – число маскирующих сигналов; m – выделяемый сигнал; $r'_{m\,лок\,max}$ — максимальное значение обостренной функции локализации для сигнала, выделяемого КИЗ; $r'_{m\,лок\,\varphi_{\max}}$ – значение обостренной функции локализации сигнала i-го КИЗ, вычисленное для направления φ_m на оцениваемый источник звука.

Влияние пространственного разнесения КИЗ на изменение условий для их выделения оценивается критерием

$$K'_{\Delta N,\Delta\tau} = K_{\Delta N\neq 0,\Delta\tau\neq 0} / K_{\Delta N=0,\Delta\tau=0},$$

где числитель и знаменатель вычисляются по формуле (1.13).

Покажем особенности применения этой модели также на примере стереопанорамы, состоящей из трех речевых источников (рис. 1.89). Демаскировка формирующих их сигналов при пространственном разнесении КИЗ

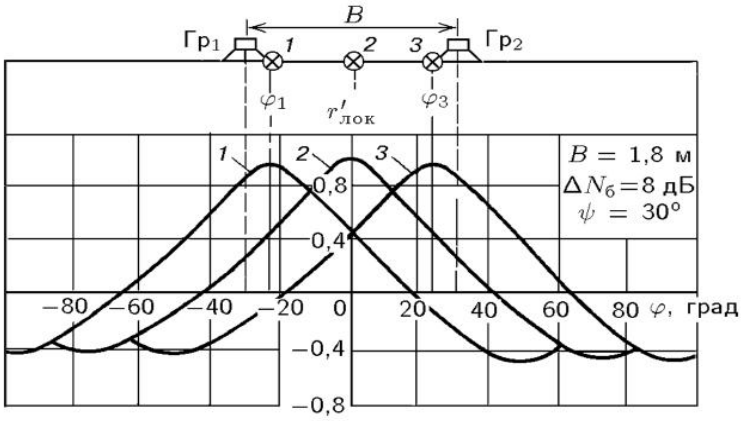


Рис. 1.89. Функции локализации сигналов, формирующих пространственную звуковую панораму, состоящую из трех речевых КИЗ: B=1,8 м, x = 0, $2\psi = 60^0$, $\Delta N_б$ = 8 дБ

сопровождается увеличением отношения $K'_{\Delta N,\Delta\tau}$ (рис. 1.90). Здесь по оси ординат отложены значения $K_{\Delta N}'$, дБ, а по оси абсцисс – разность уровней

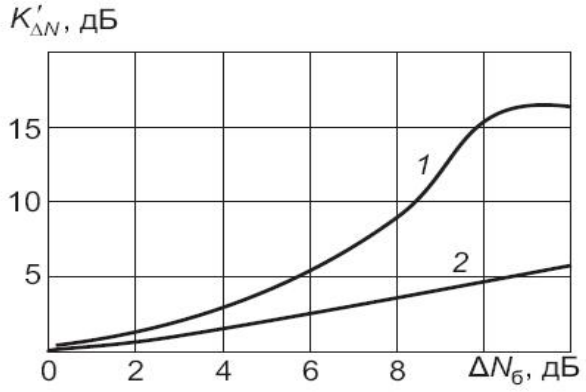


Рис. 1.90. Относительное изменение отношения сигнал/помеха для боковых (кривая 1) и центрального (кривая 2) КИЗ при интенсивностной стереофонии: В=1,8 м, x=0, 2Ψ=60^0, x=0

$\Delta N_б$, также в дБ, канальных сигналов, формирующих боковые речевые источники звука стереопанорамы для всех КИЗ, однако для боковых источников (кривая 1) это отношение растет значительно быстрее, чем для центрального КИЗ (кривая 2). Именно это создает неодинаковые условия для их выделения и воспринимается субъективно как "провал середины".

Сопоставление теории и эксперимента представлено на рис. 1.91. Вся трудность здесь состоит в том, что для получения теоретических результатов необходимо, прежде всего, связать разборчивость с величиной отношения $K'_{\Delta N}$.

Для этого воспользуемся семейством экспериментально найденных зависимостей слоговой разборчивости $S_р$,%, от уровня речи $N_р$,дБ, [1.60, 1.61], полученных при различных уровнях $N_ш$ маскирующего шума.

Зная воспринимаемый уровень выделяемого слушателем источника звука (при проведении испытаний эта величина для каждого из трех речевых источников составляла 80 дБ) и величину его слоговой разборчивости при ΔN=0, можно из упомянутых выше кривых определить уровень эквивалентного шума, оказывающего такое же маскирующее действие на полезный речевой источник, как и два остальных одновременно воспринимаемых с ним КИЗ. Теперь нетрудно определить отношение сигнал/помеха при $\Delta N = 0$:

$$10\lg K_{\Delta N=0}=(N_р)_{\Delta N=0} - (N_ш)_{\Delta N=0},$$

тогда аналогично при $\Delta N \neq 0$

$$10\lg K_{\Delta N\neq 0}=(N_р)_{\Delta N=0} - (N_ш)_{\Delta N\neq 0},$$

где величина $(N_ш)_{\Delta N\neq 0}$ может быть найдена как

$$(N_ш)_{\Delta N\neq 0} = (N_ш)_{\Delta N=0} - 10\lg(K_{\Delta N\neq 0}/ K_{\Delta N=0}) = (N_ш)_{\Delta N=0} - 10\lg K'_{\Delta N}. \quad (1.14)$$

Из изложенного следует, что разборчивость речевого КИЗ, маскируемого двумя другими речевыми источниками, может быть найдена следующим образом:

-сначала с помощью выражений (1.12) и (1.13) определяется величина коэффициента K;

-затем по формуле (1.14) вычисляется уровень эквивалентного маскирующего шума $(N_ш)_{\Delta N\neq 0}$ при известных значениях $(N_ш)_{\Delta N=0}$ и $=(N_р)_{\Delta N=0}$;

- и, наконец, из семейства кривых $S_р=f(N_р)$, выражающих зависимость слоговой разборчивости от уровня речи и шума, и найденной величине эквивалентного маскирующего шума $(N_ш)_{\Delta N\neq 0}$ определяется его слоговая разборчивость $S_р$, %.

Заметим, что известные из работ [1.60, 1.61] взаимосвязи между различными видамиартикуляции позволяют при известной слоговой разборчивости $S_р$ определить, в частности, словесную разборчивость $W_р$,%.

Полученные таким образом результаты для боковых (кривая 1) и центрального (кривая 2) речевых кажущихся источников в зависимости от $\Delta N_б$ приведены на рис. 1.91. Здесь по оси ординат отложена словесная разборчивость $W_р$,% каждого из КИЗ, а по оси абсцисс – разность уровней $\Delta N_б$ для пар сигналов, формирующих боковые КИЗ. Расхождение теории и

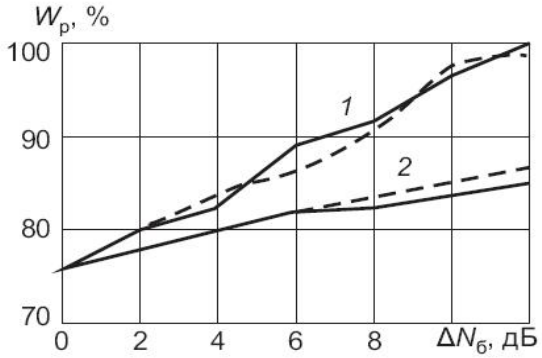


Рис. 1.91. Экспериментальные (сплошные линии) и теоретические (пунктирные линии) зависимости разборчивости речевых источников пространственной панорамы от разности уровней сигналов, формирующих боковые КИЗ: B = 1,8 м, x = 0, $2\Psi = 60^0$, x = 0

данных экспертиз (сплошные линии) не првышает здесь 10%, что подтверждает правомерность прменения данной модели.

1.14. Психоакустические модели стандартов MPEG

Наиболее известны три психоакустические модели: *NMR* (*Noise to Mask Ratio*), *PAQM* (*Perceptual Audio Quality Measure*) и *PERCEVAL* (*PERCeptual EVALution*). Самое широкое распространение получила *NMR* модель, в которой при расчете глобального (суммарного) порога маскировки учитываются абсолютный порог слышимости и явление маскировки в частотной области. Маскировка во временной области и явление демаскировки сигналов, свойственное пространственному восприятию источников звука, в *NMR* модели не учитываются, и это является ее существенным недостатком.

С помощью психоакустической модели для каждой субполосы кодирования n вычисляется отношение сигнал-маска, $SMR(n)$. Оно представляет собой выраженное в децибелах отношение энергии звукового сигнала к максимально возможному значению энергии искажений квантования в субполосе кодирования, при котором они еще маскируются полезным сигналом. Совокупность значений $SMR(n)$, вычисленных для всех субполос кодирования, образует глобальный порог маскировки, определяющий требуемое для кодирования субполосных отсчетов или соответствующих им коэффициентов МДКП минимально возможное число бит.

Психоакустическая модель 1

Она применяется в стандартах *MPEG*-1 и *MPEG*-2 для уровней компрессии *Layer* 1 и *Layer* 2. Основные этапы выполняемых в ней вычислений можно представить в виде схемы, изображенной на рис. 1.92. Здесь же даны и основные расчетные формулы.

Итак, вначале рассчитывается энергетический спектр $X(k)$ выборки входного сигнала (блок 1). Длина выборки N быстрого преобразования Фурье (БПФ) составляет $N = 512$ (*Layer* 1) или 1024 отсчета (*Layer* 2); в представленном в данном блоке выражении буква n – номер отсчета сигнала в выборке, k – номер (индекс) коэффициента БПФ. На выходе блока БПФ имеем линейчатый спектр с разрешением по частоте $\Delta F = f_Д/N$, где $f_Д$ – частота дискретизации сигнала. При $f_Д = 48$ кГц и N = 1024 получаем, что $\Delta F =$ 46,875 Гц. Перед вычислением спектра выборка отсчетов ЗС взвешивается оконной функцией Ханна $h(n)$ для уменьшения искажений, вызванных эффектом Гиббса. Вычисленный спектр нормируется: максимальной по величине спектральной компоненте присваивается уровень равный 96 дБ.

Далее (блок 2) вычисляется энергия сигнала $E_{sb}(n)$, дБ, в каждой из субполос кодирования n. Здесь: $[X_{sb(n)}(k)]_{max}$ и $SCF_{max}(n)$ –соответственно

Первичный ИКМ звуковой сигнал

1.Расчет энергетического спектра выборки звукового сигнала и его нормирование:

$$X(k) = 10\lg \left| (1/N)\sum_{n=0}^{N-1} h(n)s(n)\exp(-jkn2\pi/N) \right|^2, \text{ дБ}, \quad \text{где } k=0,1,\ldots, N/2;$$

N = 1024 или 512; $h(n)=\sqrt{8/3}\cdot 0{,}5\cdot\{1-\cos[2\cdot\pi\cdot n/N]\}$; нормирование к уровню 96 дБ.

2.Вычисление энергии сигнала выборки в субполосах кодирования:

$$E_{sb}(n) = max\{[X_{sb(n)}(k)]_{max}, 20\lg[SCF_{max}(n)\cdot 32768)-10], \text{ дБ}, \quad \text{или}$$

$$E_{sb}(n) = max[X_{sb}(n), 20\lg[SCF_{max}(n)\cdot 32768)-10], \text{дБ}, \text{ где } X_{sb}(n)=10\lg\sum_{k\in n} 10^{X(k)/10}, \text{ дБ}$$

3.Выделение локальных максимумов спектра сигнала выборки:

$$X(k) > X(k-1) \text{ и } X(k) >= X(k+1)$$

4.Формирование списка тональных компонент:

$X(k) - X(k+j) >= 7$ дБ, Обследуемая область частот $\Delta F(k+j)$, для *Layer* 2:

$$X_{tm}(k)=10\lg[\sum_{i=-1}^{i=+1} 10^{X(k+i)/10}], \text{ дБj}$$

j = -2, +2 для 2 < k < 63

j = -3,-2,+2,+3 для 63 <= k < 127

Список тональных компонент $\{X_{tm}(k)\}_{tk}$

j = -6,…,-2,+2,…,+6 для 127 <= k < 255

j = -12,…,-2,+2,…,+12 для 256 <= k< 500

5.Формирование списка не тональных (шумоподобных) компонент:

1.Исключение из исходного спектра тональных и соседних с ними компонент,
2.Разделение оставшейся части спектра на критические полосы слуха,
3.Расчет уровня энергии спектральных компонент в критических полосах слуха:

$$\{X_{nm}(k)\}_{nk}, \text{ где } X_{nm}(k_1)=10\lg[\sum_i 10^{0,1X(i)}], \text{ дБ, кроме } \forall X(i)\notin\{X_{tm}(k-1, k, k+1)\}$$

6.Прореживание спектра тональных и не тональных компонент:6

1.Исключение компонент, лежащих ниже абсолютного порога слышимости,
2.Прореживание тональных компонент с помощью окна шириной 0,5 барк,
3.Исключение компонент, удовлетворяющих условию:

$$X_{tm,nm}(i)=X_{tm,nm}(k), X_{tm,nm}(k)=0, \text{ если:}$$

$$i=\begin{cases} k & 1\le k\le 48; \\ k+(k \bmod 2) & 49\le k\le 96; \\ k+3-((k-1)\bmod 4) & 97\le k\le 232; \\ k+3-((k-1)\bmod 8) & 233\le k\le 512. \end{cases}$$

Продолжение рис. 1.92

7.Расчет коэффициентов маскировки и индивидуальных кривых маскировки для тональных и нетональных (шумоподобных) маскирующих компонент спектра сигнала выборки:

1.Коэффициенты маскировки тональных компонент:

$$K_{M1}[z(i)] = -0{,}275z(j) - 6{,}025, \text{ дБ};$$

2.Коэффициенты шумоподобных компонент:

$$K_{M2}[z(i)] = -0{,}175z(j) - 2{,}025, \text{ дБ};$$

3.Кривые индивидуальной маскировки:

$$M[z(i),z(j)] = \begin{cases} 17(\Delta z(i,j)+1) - \{0{,}4X([z(j)] + 6\}, \text{ дБ}, & \text{для } -3 <= \Delta z(i,j) < -1; \\ \{0{,}4\cdot X[z(j)] + 6\}\cdot \Delta z(i,j), \text{ дБ}, & -1 <= \Delta z(i,j) < 0, \\ -17\cdot\Delta z(i,j), \text{ дБ}, & 0 <= \Delta z(i,j) < +1; \\ -(\Delta z(i,j) - 1)\cdot\{17 - 0{,}15\cdot X[z(j)]\} - 17, \text{ дБ}, & +1 <= \Delta z(i,j) < +8; \end{cases}$$

барк, причем $\Delta z(i,j)=z(i)-z(j)$.

8.Расчет порогов маскировки для тональных и шумоподобных компонент спектра сигнала выборки:

1.Тональные компоненты: $N_{tm}[z(i)] = X_{tm}[z(i)] + K_{M1}[z(i)] + M[z(i),z(j)]$, дБ;

2. Шумоподобные компоненты: $N_{nm}[z(i)] = X_{nm}[z(i)] + K_{M2}[z(i)] + M[z(i),z(j)]$, дБ

9.Расчет глобального порога маскировки и отношения сигнал-маска *SMR* в суб-полосах кодирования:

1.Расчет кривой глобального порога маскировки:

$$N_{ПМ}(i)=10\lg[10^{N_{апс}[z(i)]/10} + \sum_{l=1}^{L} 10^{N_{tm}[z(i),z(j)]/10} + \sum_{p=1}^{M} 10^{N_{nm}[z(i),z(j)]/10}], \text{ дБ},$$

2.Расчет минимального порога маскировки в субполосе кодирования:

$$N_{ПМ}(n)=[N_{ПМ}(i)]_{min}, \text{ дБ},$$

2.Расчет отношения сигнал-маска в субполосе кодирования n:

$$SMR(n) = E_{sb}(n) - N_{ПМ}(n), \text{ дБ}.$$

Значения $SMR(n)$ в полосах кодирования

Рис. 1.92. Последовательность вычислений в психоакустической модели 1 стандартов *MPEG*

уровень энергии максимальной по величине спектральной компоненты и наибольшее из трех значение масштабного коэффициента в субполосе кодирования n; число «10» корректирует разность между пиковым и средним значениями уровня сигнала; $X_{sb}(n)$ – выраженная в дБ суммарная энергия всех спектральных компонент в субполосе кодирования n. Заметим, что в

блоке 2 даны два выражения для расчета $E_{sb}(n)$. Более точным из них является нижнее выражение.

В блоке 3 вычисляются локальные максимумы энергетического спектра сигнала выборки. Спектральная компонента $X(k)$ считается локальным максимумом, если она по величине больше предшествующей $X(k\text{-}1)$, но не менее следующей $X(k+1)$ компоненты. Затем все спектральные составляющие сигнала выборки разделяются на тональные $X_{tm}(k)$ (блок 4) и нетональные (шумопобные) $X_{nm}(k_1)$ (блок 5) компоненты.

Для выделения тональных компонент (блок 4) исследуется область частот вокруг каждой спектральной компоненты, являющейся локальным максимумом. Соответствующая ему спектральная составляющая включается в список тональных компонент $\{X_{tm}(k)\}_{tk}$, если в обследованной вокруг нее области частот она превышает любую компоненту, исключая две соседние с ней, не менее чем на 7 дБ. Значения j определяют границы обследуемых областей частот $\Delta F(k+j)$. Эти области расширяются с повышением частоты, то есть с ростом индекса k спектральной компоненты $X(k)$. При расчете уровня тональной компоненты $X_{tm}(k)$ учитываются энергии двух соседних с каждой из них компонент $X(k-1)$ и $X(k+1)$. В результате формируется список тональных компонент $\{X_{tm}(k)\}_{tk}$. Как правило, число тональных компонент сравнительно невелико.

После этого формируется список нетональных (шумоподобных) компонент $\{X_{nm}(k)\}_{nk}$ (блок 5). Для этого из исходного спектра сигнала выборки исключаются тональные и соседние с ними компоненты, уже учтенные ранее. Затем спектр оставшихся компонент разделяется на полосы частот равные критическим полосам слуха. Границы критических полос в стандартах *MPEG* заданы таблично. В каждой из них вычисляется суммарная энергия шумоподобных компонент. Далее все шумоподобные компоненты внутри критической полосы слуха замещаются одной компонентой $X_{nm}(k_1)$ с равной энергией и расположенной в центре соответствующей критической полосы слуха, разумеется, с учетом дискретности ΔF спектра сигнала выборки. Сформированный таким образом список $\{X_{nm}(k)\}_{nk}$ будем называть шумоподобными (нетональными) компонентами. Он содержит не более 24-х компонент, соответственно по одной в каждой критической полосе слуха.

Разделение исходного спектра сигнала выборки на тональные и нетональные (шумоподобные) компоненты необходимо, так как значения коэффициентов маскировки для них имеют разные величины. Они оценивают маскировку внутри критической полосы слуха (*intra-band-masking*). В качестве примера на рис. 1.36,*б* показаны значения коэффициентов маскировки в дБ, в функции от высоты тона z, в барках, для двух ситуаций:

-тон маскирует шум (*tone masking noise*), последний равномерно охватывает этот тон с двух сторон и имеет полосу частот равную критической полосе слуха (рис. 1.36,*б* нижняя прямая); в стандартах *MPEG* зависимость

коэффициента маскировки для этой ситуации описывается формулой, представленной в пункте 1 блока 7;

-шум, имеющий полосу частот равную критической полосе слуха, маскирует тон (*noise masking tone*), расположенный внутри него (рис. 1.36,*б*, верхняя линия). В этой ситуация значение коэффициента маскировки в функции от высоты рассчитывается по формуле, приведенной в пункте 2 блока 7. Оба этих выражения заимствованы из стандарта *MPEG*-1. Напомним, что частота F (Гц) и высота тона z (барк) связаны эмпирической зависимостью вида:

$$z = 13\cdot\mathrm{arctg}(0{,}0076\cdot F) + 3{,}5\cdot\mathrm{arctg}[(F/7500)]^2, \text{ барк.}$$

Нетрудно видеть (рис. 1.36), что значения коэффициентов маскировки в обоих случаях падают с ростом высоты тона z. В общем случае для оценки маскировки внутри критической полосы слуха используется понятие коэффициента (индекса) тональности α и выражение вида:

$$K_{\mathrm{M}(i)} = -[\alpha\cdot(14{,}5 + i) + (1{,}0 - \alpha)\cdot 5{,}5], \text{ дБ.}$$

Для чистого тона индекс тональности α равен 1 и значение коэффициента маскировки меняется от -14,5 дБ для первой критической полосы слуха (i = 1) до значения – (14,5 +24) = - 38,5 дБ (при i = 24), (*tone masking noise*). Для шумоподобного сигнала (*noise masking tone*) значение коэффициента маскировки $K_{\mathrm{M}(i)}$ в первом приближении равно -5,5 дБ и не зависит от его положения на шкале Барков. Однако более точные данные для этой ситуации получены Э.Цвикером, который предложил использовать для расчета коэффициента маскировки в ситуации, когда шум маскирует тон выражение вида:

$$K_{\mathrm{M2}} = -2{,}2 - 2{,}05\cdot\mathrm{arctg}(F/4) - 0{,}75\cdot\mathrm{arctg}(F^2/2{,}56^2), \text{ дБ.}$$

Здесь F выражено в кГц, а K_{M2} – в дБ. Учитывая различие в маскировке шумоподобными и тональными компонентами, можно обобщенное выражение для расчета коэффициента маскировки записать иначе:

$$K_{\mathrm{M}(i)} = \alpha(i)\cdot(14{,}5 + i) + (1{,}0 - \alpha)\cdot K_{\mathrm{M2}}(i), \text{ дБ,}$$

где i – номер критической полосы слуха (i = 1,2,…,24) или высота тона, $\alpha(i)$ – индекс тональности, изменяющийся от нуля (шумоподобный сигнал) до 1, когда маскирующим сигналом является чистый тон.

Маскировка вне критической полосы *слуха* (*extra-band-masking*) одинакова как для тональных, так и для нетональных компонент. Она оценивается с помощью *индивидуальных кривых маскировки*, учитывающих избирательные свойства базилярной мембраны слухового анализатора и взаимное маскирующее действие соседних спектральных компонент. Однако до

построения этих кривых спектры тональных и нетональных компонент прореживаются (блок 6). Прежде всего, исключаются из рассмотрения все тональные $X_{tm}(k)$ и нетональные $X_{nm}(k)$ компоненты, лежащие ниже абсолютного порога слышимости. Кроме того, тональные компоненты дополнительно прореживаются с помощью окна шириной 0,5 барка. Если в окно попало две тональных компоненты, то та из них, которая имеет меньший уровень, выбрасывается. Вообще говоря, ширина этого окна при прореживании не одинакова в разных психоакустических моделях.

После прореживания формируется новая сетка спектральных компонент. При этом в первых трех субполосах кодирования (0…2250 Гц) учитываются все спектральные компоненты; в следующих трех субполосах (2250..4500 Гц) – каждая вторая; в последующих трех субполосах (4500…6750 Гц) – уже каждая четвертая; и наконец, в оставшихся 20 субполосах – лишь каждая восьмая спектральная составляющая (рис. 1.92, блок 6, пункт 3). В итоге, если верхняя частота ЗС ограничена значением 22500 Гц, то после такого прореживания получаем спектр, состоящий в общей сложности из 126 спектральных компонент. Напомним, что исходный спектр содержал 512 спектральных составляющих.

В блоке 7 рассчитываются коэффициенты маскировки для тональных $K_{M1}[z(i)]$ и не тональных $K_{M2}[z(i)]$ компонент, а также индивидуальные кривые маскировки $M[z(i),z(j)]$ для каждой из них. Здесь $z(i)$ – высота тона маскирующей компоненты (тональной или шумоподобной); $z(j)$ – высота тона маскируемой компоненты; $\Delta z(i,j)$ – разность высот тона маскирующей $z(i)$ и маскируемой $z(j)$ компонент, также в барках. Семейство индивидуальных кривых маскировки представлено на рис. 1.37,*а*. По оси ординат отложены значения относительного порога слышимости (порога маскировки) М в дБ, вычисленные с помощью выражений, приведенных в пункте 3 блока 7, рис. 1.92. По оси абсцисс – разность высот тона $\Delta z = z(i) - z(j)$ маскируемой $z(j)$ и маскирующей $z(i)$ компонент, в барках. Параметром представленных кривых является значение уровня маскирующей компоненты в дБ. Каждый спад кривой маскировки в сторону верхних и нижних частот аппроксимирован здесь двумя отрезками прямых линий.

В блоке 8 рассчитываются пороги маскировки $N_{tm}[z(i),z(j)]$ и $N_{nm}[z(i),z(j)]$ для каждой тональной $X_{tm}[z(i)]$ и шумоподобной $X_{nm}[z(i)]$ компонент.

И наконец, в блоке 9 психоакустической модели вычисляются: кривая глобального порога маскировки $N_{ПМ}(i)$ для выборки ЗС путем суммирования порогов маскировки тональных и шумоподобных компонент (пункт 1); минимальное значение порога маскировки в $N_{ПМ}(n)$ в каждой из субполос кодирования n (пункт 2), а после этого рассчитывается уже отношение сигнал-маска $SMR(n)$ для каждой из субполос кодирования n (пункт 3 блока 9). Заметим, что здесь $N_{АПС}[z(i)]$ – выраженное в дБ значение абсолютного порога слышимости спектральной компоненты с высотой тона $z(i)$.

В психоакустической модели 1 исповедуется принцип аддитивности, (взаимонезависимости) действия на орган слуха спектральных компонент при их одновременном предъявлении. Напомним, что величина *SMR*(*n*) представляет собой выраженное в дБ отношение энергии полезного сигнала к максимально допустимому значению энергии искажений квантования в субполосе кодирования *n*, при котором они еще маскируются полезным сигналом.

Психоакустическая модель 2

В стандартах *MPEG ISO/IEC* 11172-3 и 13818-3 она использована в алгоритме компрессии *Layer* 3, а также ее модификация – в стандартах *MPEG ISO/IEC* 13818-7 *AAC* и 14496-3.

Ее характерными особенностями являются:

-разделение спектра выборки звукового сигнала на полосы психоакустического анализа *b*, в которых и происходят вычисления (это не критические полосы слуха); при этом имеем 62 полосы анализа для частоты дискретизации 48 кГц, 63 полосы анализа для частоты дискретизации 44.1 кГц и 59 полос анализа для частоты дискретизации 32 кГц;

-явления временной маскировки (предмаскировки и постмаскировки), а также пространственной демаскировки сигналов здесь также не учитываются;

-все расчеты одновременно выполняются как для длинных (N = 1024), так и для трех коротких (N = 256) выборок звукового сигнала.

Расчет отношения сигнал-маска *SMR(n)* в полосах кодирования *n* состоит здесь из процедур, указанных в блоках на рис. 1.93. Здесь же приведены и основные расчетные формулы. Назовем эти процедуры.

1.Вычисление БПФ, в результате которого для каждой спектральной компоненты $Y_w = \alpha_w + j\beta_w$ сигнала выборки вычисляются ее амплитуда $r_w = /Y_w/$ и фаза φ_w (блок 1), где w -индекс (номер) спектральной компоненты. Перед выполнением ортогонального преобразования сигнал выборки *x(n)*, как и в психоакустической модели 1, взвешивается оконной функцией Ханна *h(n)*.

2.Вычисление предсказанных значений амплитуды $\hat{r}_w$ и фазы $\hat{\varphi}_w$ для каждой спектральной компоненты w сигнала текущей выборки; для этой цели в памяти кодера *MPEG* должны храниться массивы значений модулей и фаз спектральных составляющих двух блоков *t - 1* и *t - 2*, предшествующих текущему блоку *t* (рис. 1.93, блок 2).

3.Расчет меры непредсказуемости C_w для каждой спектральной компоненты w текущей выборки (блок 3 на рис. 1.93). Эта мера учитывает наличие корреляционной связи между соответствующими спектральными компонентами текущего *t* и предшествующими *t - 1* и *t - 2* выборками. На

Первичный ИКМ сигнал

1.Расчет спектра выборки звукового сигнала:

$$Y_w = \frac{1}{N}\sum_{n=0}^{N-1} s(n)\cdot h(n)\cdot \exp(-j2\pi wn/N)\text{, дБ,} \qquad w = 0,1,\ldots, N-1;$$

$$N=1024 \text{ или } N=256;\ h(n)= 0,5\cdot\{1-\cos[2\cdot\pi(n-0,5)/N]\};$$

$$Y_w = \alpha_w + j\cdot\beta_w;\ r_w = /Y_w/;\ \varphi_w=argtg(Y_w)$$

2.Вычисление предсказанных значений амплитуды r_w и фазы φ_w спектральных составляющих текущей выборки звукового сигнала (ЗС):

$$\hat{r}_w = 2\cdot r_w(t-1) - r_w(t-2); \quad \hat{\varphi}_w = 2\cdot\varphi_w(t-1) - \varphi_w(t-2)$$

3.Расчет меры непредсказуемости спектральных компонент текущей выборки ЗС:

$$C_w = \begin{cases} c_{l(w)} & 0 \le w < 6 \\ c_{s((w+2)DIV\ 4)} & 6 \le w < 206 \\ 0.4 & w \ge 206 \end{cases};$$

$$c_w = \frac{\sqrt{(r_w\cdot\cos(\varphi_w) - \hat{r}_w\cdot\cos(\hat{\varphi}_w))^2 + (r_w\cdot\sin(\varphi_w) - \hat{r}_w\cdot\sin(\hat{\varphi}_w))^2}}{r_w + \left|\hat{r}_w\right|}$$

4.Вычисление энергии сигнала и взвешенного значения меры непредсказуемости в полосах психоакустического анализа:

$$e_b = \sum_{w=wlow_b}^{whigh_b} r_w^2; \qquad c_b = \sum_{w=wlow_b}^{whigh_b} r_w^2\cdot C_w$$

5.Свертывание энергии сигнала и взвешенного значения меры непредсказуемости с развертывающей функцией:

$$ec_b = \sum_{bb=1}^{b\max} e_{bb}\cdot M(bval_{bb}, bval_b); \qquad ct_b = \sum_{bb=1}^{b\max} c_{bb}\cdot M(bval_{bb}, bval_b)$$

$$M(i,j) = \begin{cases} 0 & при \quad tmpy < -100 \\ 10^{(tmpz+tmpy)/10} & при \quad tmpy \ge -100 \end{cases};$$

$$tmpz = 8\cdot\min\left\{(tmpx-0.5)^2 - 2(tmpx-0.5),\ 0\right\}$$

$$tmpy = 15.811389 + 7.5(tmpx+0.474) - 17.5\left\{1.0 + (tmpx+0.474)^2\right\}^{1/2}$$

$$tmpx = \begin{cases} 3.0(j-i) & при \quad j \ge i \\ 1.5(j-i) & при \quad j < i \end{cases}$$

6.Расчет коэффициента Хаоса и индекса тональности в полосах психоакустического анализа b:

$$cb_b = ct_b / ec_b \qquad \alpha_b = -0.299 - 0.43\ln(cb_b)$$

Рис.1.93.Последовательность вычислений в психоакустической модели 2 стандартов MPEG

Продолжение вычислений в модели 2

7.Расчет отношения сигнал-шум SNR_b в полосах психоакустического анализа b:

$SNR_b = \max(minval_b, SNR^*_b)$,дБ; $SNR^*_b = \alpha_b \cdot TMN_b + (1-\alpha_b)\cdot NMT_b$, дБ ;

$$bc_b = 10^{-SNR_b/10}$$

8. Расчет энергии шума на пороге его слышимости, приходящейся на один коэффициент МДКП в полосе психоакустического анализа *b*:

1.Модель 2: $thr_w = \max(nb_w, athr_w)$, где $nb_w = \frac{nb_b}{whigh_b - wlow_b + 1}$, $nb_b = en_b \cdot bc_b$,

2.*Layer* 3, в целом для полосы анализа *b*:

а) глобальный порог маскировки для длинных блоков (*N=1024*):

$$thr_b = \max\left(athr_b, \min(nb_b, nb_b(t-1), nb_b(t-2))\right)$$

$$nb_b(t-1) = \eta_1 \cdot nb_b; \quad nb_b(t-2) = \eta_2 \cdot nb_b; \quad \eta_1 = 2; \eta_2 = 16$$

б) глобальный порог маскировки для коротоких блоков (*N=256*):

$$thr_b = \max\left(athr_b, nb_b\right)$$

9. Расчет глобального порога маскировки (допустимой энергии шума) в полосах кодирования:

$$thr_n = \begin{cases} \sum_{w=wlow_n}^{whigh_n} thr_w & \text{при } \mathrm{width_n} = 1 \\ \min(thr_{wlow_n}, \ldots\ldots, thr_{whigh_n}) \cdot (whigh_n - wlow_n + 1) & \text{при } \mathrm{width_n} = 0 \end{cases}$$

10.Расчет энергии звукового сигнала в полосах кодирования *n*:

$$e_n = \sum_{w=wlow_n}^{whigh_n} r_w^2$$

11.Расчет отношения сигнал-маска SNR_n в полосах кодирования *n*:

$$SMR_n = 10 \cdot \lg\left(\frac{e_n}{thr_n}\right);$$

Процедура вычислений значений e_n и thr_n для *Layer* 3:

а) энергия звукового сигнала:

$$e_n = w1 \cdot e_{b_u} + \sum_{b=b_u+1}^{b=b_0-1} e_b + w2 \cdot e_{b_0}$$

б) глобальный порог маскировки:

$$thr_n = w1 \cdot thr_{b_u} + \sum_{b=b_u+1}^{b=b_0-1} thr_b + w2 \cdot thr_{b_0}$$

$w1, w2, b_u, b_o$ - табличные данные

Выход психоакустической модели 2 стандартов *MPEG*

Рис.1.93.Последовательность вычислений в психоакустической модели 2 стандартов *MPEG* (Продолжение)

основании меры непредсказуемости делается вывод о степени близости сиг нала в полосе психоакустического анализа b к тону или к шуму, маскирующие свойства которых, как известно, различны. В модифицированной психоакустической модели 2 массив значений C_w включает следующие данные: для спектральных компонент с индексами $0 \le w < 6$ значения C_w берутся из длинных блоков выборки; для $6 \le w < 206$ – из второго короткого блока и для спектральных компонент с индексами $w \ge 206$ величина меры непредсказуемости принимается равной значению 0,4; знак *DIV* в формуле для расчета значений меры непредсказуемости обозначает целочисленное деление с округлением результата в сторону $-\infty$. Значение меры непредсказуемости c_w для каждой спектральной компоненты вычисляется (независимо от длины блока) по одной и той же формуле, представленной справа (блок 3).

4.В блоке 4 рассчитывается энергия e_b и взвешенное значение меры непредсказуемости c_b текущей выборки звукового сигнала в каждой полосе психоакустического анализа b. В приведенных здесь выражениях буквами $wlow_b$ и $whigh_b$ обозначены соответственно нижний и верхний индексы w спектральных компонент в полосе психоакустического анализа b, где b номер (индекс) полосы анализа. Эти значения заданы в стандартах *MPEG* отдельными таблицами для каждой частоты дискретизации.

5. В блоке 5, прежде всего, рассчитывается так называемая развертывающая функция $M(i,j)$, представляющая собой индивидуальную кривую маскировки, учитывающую избирательные свойства базилярной мембраны уха. В формуле для расчета развертывающей функции приняты следующие обозначения: $i=bval_{bb}$ – значение высоты тона в барках для развертываемого сигнала, $j=bval_b$ – высота тона сигнала, который развертывается в полосу анализа i, также в барках; величины *tmpz*, *tmpy* и *tmpx* являются временными переменными. Семейство данных развертывающих функций представлено на рис. 1.37,б. Здесь по оси ординат отложены значения развертывающих функций M, в дБ, а по оси абсцисс – разность высот тона Δz маскируемой и маскирующей компонент, в барках. Параметром представленных кривых является значение уровня N, дБ, маскирующей компоненты. Заметим, что учитываются только те значения развертывающей функции, которые превышают величину 10^{-6}, в противном случае они принимаются равными нулю. Развертывающие функции, как и индивидуальные кривые маскировки в психоакустической модели 1, строятся в шкале высот тона, измеряемой в барках.

Далее рассчитываются свертки (ec_b и ct_b) энергий сигнала e_b и взвешенных значений меры непредсказуемости c_b с развертывающей функцией $M(i,j)$ (она обозначена в стандарте как $sprdgnf(bval_{bb}, bval_b)$) с целью учета влияния соседних полос психоакустического анализа (блок 5).

6.Расчет отношения сигнал-шум SNR_b в полосах психоакустического анализа b (блок 7 на рис. 1.93) выполняется с использованием индекса то-

нальности α_b и значений коэффициентов маскировки TMN_b и NMT_b . Здесь $SNR*_b$ – наименьшее значение отношения энергий полезного сигнала и шума, выраженное в дБ, при котором этот шум еще маскируется полезным сигналом для полосы анализа b; TMN_b- разность между уровнями тона и шума, в дБ, для ситуации, когда тон маскирует шум и уровень шума соответствует его порогу слышимости, для всех полос анализа b эта величина принята равной 29 дБ (стандарт *ISO/IEC* 11172-3) и 18 дБ (стандарт *ISO/IEC* 13818-3); NMT_b – разность между уровнями шума и тона в дБ для ситуации, когда шум маскирует тон и уровень тона соответствует его порогу слышимости, для всех полос анализа b NMT_b также постоянная величина равная 6 дБ (стандарты *ISO/IEC* 11172-3 *Layer* 3 и *ISO/IEC* 13818-3).

Очевидно, чем меньше отношение $SNR*_b$, тем большим может быть допустимый уровень шума в полосе анализа b. Нижней границей отношения $SNR*_b$ служит табличная величина $minval_b$, представляющая собой поправочный коэффициент, который отличается от нуля лишь на самых нижних частотах.

Индекс тональности α_b рассчитывается с использованием коэффициента Хаоса cb_b равного отношению ct_b и ec_b (блок 6). Значения α_b ограничиваются пределами $0 < \alpha_b < 1$. Если расчеты дают значения большие 1, то его значение принимается равным 1, значения меньшие нуля заменяются при вычислениях нулями.

7.Расчет максимально-допустимой энергии шума thr_b (глобального порога маскировки) в полосе психоакустического анализа b, при которой он еще маскируется полезным сигналом (блок 8). В пункте 1 блока 8 даны расчетные формулы, используемые в психоакустической модели 2. Здесь nb_b – максимальное значение энергии шума в полосе анализа b, при котором он еще маскируется полезным сигналом; nb_w – то же самое, но приходящееся на один коэффициент МДКП сигнала выборки в полосе психоакустического анализа b; $athr_w$ – значение абсолютного порога слышимости для спектральной компоненты с индексом w в линейных единицах; $wlow_b$ и $whigh_b$ – соответственно нижняя и верхняя границы полосы психоакустического анализа b. Отметим, что для всех коэффициентов МДКП в пределах одной полосы анализа b эта величина принимается одинаковой, независимо от значений их амплитуд.

Процедура вычисления глобального порога маскировки thr_b в *Layer* 3 изменена (пункт 2, блок 8). В Layer 3 расчет допустимой энергии шума ведется в целом для полосы психоакустического анализа b. Значение глобального порога маскировки thr_b определяется здесь путем сравнения величин $athr_b$, nb_b, $nb_b(t\text{-}1) = \eta_1 \cdot nb_b$ и $nb_b(t\text{-}2) = \eta_2 \cdot nb_b$, где $athr_b$ – энергия шума в полосе анализа b, соответствующая абсолютному порогу слышимости, когда нет мешающих звуков, она задана в табличной форме для каж-

дой полосы анализа b; $nb_b(t\text{-}1)$ и $nb_b(t\text{-}2)$ – соответственно значения энергии шума на пороге слышимости, вычисленные в полосе анализа b для двух предшествующих выборок звукового сигнала, причем η_1=2 и η_2=16 – поправочные коэффициенты, их значения получены эмпирическим путем.

8.Вычисление максимально-допустимой энергии шума (глобального порога маскировки) thr_n в субполосах кодирования n (блок 9). При переходе от полос психоакустического анализа b (число этих полос зависит от значения частоты дискретизации $f_Д$ и может быть равно в *Layer* 3 соответственно 59, 62 или 63) к полосам кодирования n (их число равно 32) вводится понятие так называемой психоакустически узкой субполосы (ширина которой меньше, чем приблизительно 1/3 критической полосы) и психоакустически широкой субполосы. Ширина этих полос обозначена как $width_n$, где n – номер субполосы кодирования. При этом для психоакустически узкой полосы кодирования имеем $width_n = 1$, а для психоакустически широкой значение $width_n = 0$. В формуле для вычисления допустимой энергии шума thr_n в субполосе кодирования n (блок 9 на рис. 1.93) обозначения $wlow_n$ и $whigh_n$ представляют собой соответственно нижний и верхний индексы спектральных компонент выборки ЗС в субполосе кодирования n.

9. Вычисление энергии сигнала e_n в полосах кодирования n (блок 10). Здесь $wlow_n$ и $whigh_n$ – значения соответственно нижнего и верхнего индексов спектральных коэффициентов выборки звукового сигнала в субполосе кодирования n.

10. И, наконец, в блоке 11 рассчитывается отношение энергии полезного сигнала e_n к допустимому значению энергии шума thr_n (или так называемое отношение сигнал-маска SMR_n), передаваемое кодеру *MPEG* для каждой из субполос кодирования n.

В отличие от модели 1 здесь изменены процедуры вычислений энергий полезного сигнала e_n и глобального порога маскировки thr_n (блок 11) в субполосах кодирования. При этом значения величин $w1, w2, b_u, b_o$, берутся из соответствующих таблиц стандартов *MPEG*, представленных отдельно для каждой частоты дискретизации ЗС и соответственно для длинной или короткой выборок сигнала.

Заметим, что для коротких выборок ЗС (N = 256) в стандарте *MPEG ISO/IEC* 11172-3 *Layer* 3 принят несколько упрощенный вариант вычисления отношения сигнал-маска. Вычисление глобального порога маскировки в каждой полосе психоакустического анализа b выполняется аналогично тому, как это делалось ранее для длинных блоков. Однако, при вычислении допустимой энергии шума $bc_b = 10^{-SNR_b/10}$ (блок 7 на рис. 1.93) значения SNR_b (в дБ) для коротких выборок берутся непосредственно из таблиц стандарта *MPEG*, а не рассчитываются с помощью индекса тональности,

как это делается для длинных выборок. При этом максимально-допустимое значение энергии шума thr_b в полосе психоакустического анализа b для коротких выборок определяется сравнением значения $nb_b=en_b \cdot bc_b$ с величиной абсолютного порога слышимости $athr_b$ и выбора наибольшего из этих двух значений $thr_b = \max(athr_b, nb_b)$ (пункт 2,б блока 8 на рис. 1.93). Дальнейшая процедура вычислений, выполняемых для коротких выборок, не отличается от ранее изложенной. Глобальный порог маскировки рассчитывается отдельно для каждой из трех коротких выборок.

Заметим, что в стандартах *MPEG ISO/IEC* 13818-7 и 14496-3 применяется модифицированная психоакустическая модель 2. Введенные в ней изменения носят частный характер, поэтому отдельно здесь не рассматриваются.

1.14. Психоакустическая модель стандарта ATSC Dolby AC-3

В кодере *Dolby AC*-3 кодированию подвергаются не отсчеты ЗС, а коэффициенты МДКП. При этом каждый коэффициент МДКП представляется в формате с плавающей запятой двумя значениями: экспонентой (или порядком) и мантиссой:

$$X_D[k]=A[k]\cdot 2^{-B[k]},$$

где A[k] и B[k] –соответственно мантисса и порядок k-того коэффициента преобразования. Порядок равен числу нулей перед первой единицей двоичного представления коэффициента МДКП. Он является по сути дела его масштабным коэффициентом (или нормирующим множителем). Знак коэффициента МДКП учитывается только при кодировании мантиссы.

Массивом входных данных для блока психоакустической модели (рис.1.94) являются значения порядков B[k].

1.Преобразование массива значений порядков коэффициентов МДКП и формирование полос психоакустического анализа. Значение порядка B[k] каждого коэффициента МДКП преобразуется в значение PSD[k] для новой шкалы, содержащей 3072 градации, по формуле:

$$PSD[k]=3072- (B[k]\cdot 128).$$

В результате формируется новый массив значений порядков коэффициентов МДКП (рис.1.95).

Дискретность изменения величины порядка коэффициента МДКП как, уже было составляет 6 дБ, а наибольшее его значение равно 24. Поэтому полный динамический диапазон сигнала в системе *Dolby AC*-3 равен $6\cdot 24 = 144$ дБ.

Первичный ИКМ сигнал

1.Расчет модифицированного дискретного косинусного преобразования (МДКП) для выборки звукового сигнала и формирование полос психоакустического анализа

$$X_D[k] = \frac{-2}{N}\sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi}{4N}(2n+1)(2k+1)+\frac{\pi}{4}(2k+1)(1+\alpha)\right)$$

, при 0 <= k < N/2,

α = -1 для первого сегмента "короткого" преобразования;
= 0 для длинного преобразования;
= +1 для второго сегмента "короткого" преобразования;

$X_D[k]=A[k]\cdot 2^{-B[k]}$

2.Расчет энергии звукового сигнала в полосах психоакустического анализа:

log(a+b)=max[log(a),log(b)]+log[1+exp(d)], где

log(a) , log(b) – значения порядков соседних по частоте (в пределах одной полосы психоакустического анализа) коэффициентов МДКП;

2.Формирование обобщенной кривой маскировки:3

1.Формирование обобщенной кривой маскировки и ее аппроксимация двумя отрезками прямых линий, учитывающих маскировку лишь в сторону верхних частот (*Fast Upwards Masking*-быстро затухающая прямая и *Slow Upwards Masking*-медленно затухающая прямая);

2.Кодирование параметров обобщенной кривой маскировки 4-мя параметрами:

Slow Decay - крутизна медленно затухающего сегмента (-0,70 до -0,98 дБ/полосу анализа);

Slow Gain- вертикальное смещение медленно затухающего сегмента от уровня маскирующей компоненты сигнала (от -49 до -63 дБ);

Fast Decay- крутизна быстро затухающего сегмента (от -2,95 до -5,77 дБ/полосу анализа);

Fast Gain - вертикальное смещение быстро затухающего сегмента прямой от максимального уровня спектральной компоненты маскера (от -6 до -48 дБ);

3.Синтез обобщенной кривой маскировки:

$$x_0(k) = [x_0(k) - d_0(k)] \oplus [E_C(k)-g_0(k)]$$
$$x_1(k)=[x_1(k)-d_0(k)] \oplus [E_C(k)-g_1(k)]$$
$$E_M(k)= \max(x_0,x_1)$$

где $E_C(k)$- энергия звукового сигнала в полосе психоакустического анализа k; $d_0(k)$ и $d_1(k)$ - крутизна наклона для быстро и медленно затухающего сегментов (*Fast Decay и Slow Decay)* обобщенной кривой маскировки; $x_0(k)$ и $x_1(k)$ - вертикальное смещение сегментов от максимального уровня к-той спектральной компоненты сигнала, соответственно для быстро и медленно затухающего сегментов (*Fast Gain[ch] и Slow Gain);* $\oplus$ - оператор "*log-addition*" (логарифмическое сложение), в алгоритме Dolby AC-3 этот оператор заменен на оператор *max*.

3.Расчет кривой глобального порога маскировки и отношения сигнал-маска MR_n (*k*):

1.Массив значений допустимых энергий шумов квантования для каждого коэффициента МДКП сигнала выборки:

2.Массив значений сигнал-маска SMR_n для каждой субполосы кодирования n

Выход психоакустической модели системы *Dolby AC*-3

Рис. 1.94. Последовательность вычислений в психоакустической модели алгоритма компрессии системы *Dolby AC*-3

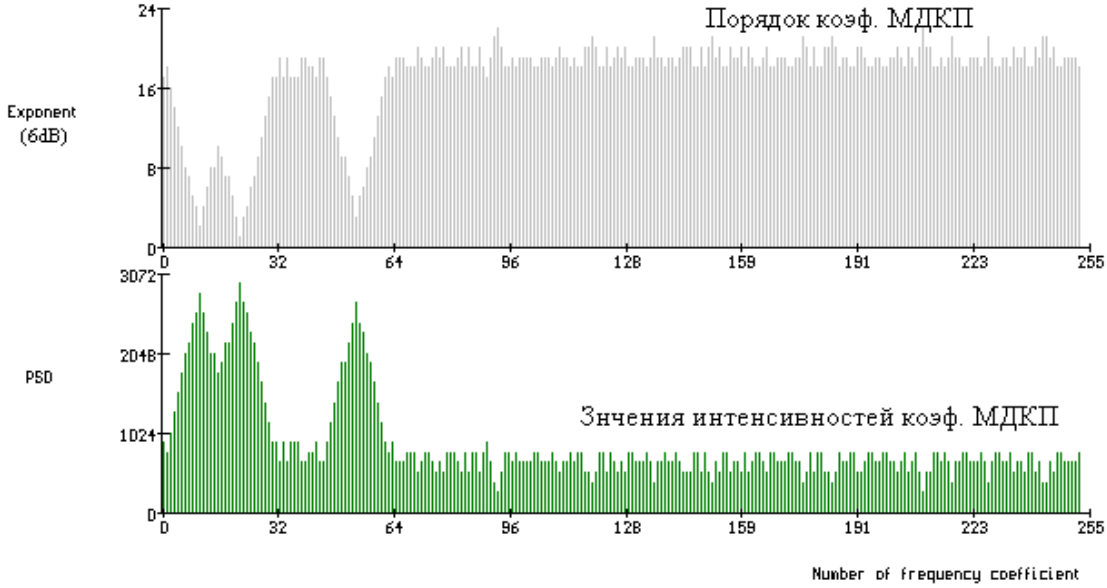


Рис. 1.95. Расчет интенсивностей порядков коэффициентов МДКП

Множитель 128 в данной формуле позволяет уменьшить дискретность грубой шкалы до величины шага равного 6:128 = 0,046875 дБ, что лежит уже существенно ниже порога различимости слуха по амплитуде. Полученная таким образом новая шкала значений, изменяющихся в диапазоне от 0 до 3072 с шагом 0,046875 дБ, если эти величины выражать в числе дискретных ступеней, принимается далее в качестве основной. Все значения параметров в стандарте *Dolby AC*-3 выражены в абсолютных единицах этой новой шкалы.

Вычисления в психоакустической модели выполняются в так называемых полосах психоакустического анализа. Они не одинаковы по ширине. Различие между структурой полос психоакустического анализа, принятой в системе *Dolby AC*-3, и критическими полосами слуха иллюстрирует рис.1.96. По оси ординат отложены значения ширины полос анализа и критических полос слуха, а по оси абсцисс – значения средней частоты сигнала. При этом ступенчатая кривая соответствует полосам анализа в кодере *Dolby AC*-3, сплошная кривая (*Critical Bandwidth*) – критическим полосам слуха, а пунктирная линия – полосам, ширина которых соответствует половине ширины критических полос слуха (0,5 *Critical Bandwidth*).

До частоты 2531 Гц ширина полос психоакустического анализа одинакова и выбрана так, что в каждую из них попадает только один коэффициент МДКП. Например, при частоте дискретизации $f_Д$ = 48 кГц и длине выборки N = 512 отсчетам ЗС ширина этих полос составляет 93,75Гц. До частоты 2531 Гц имеем в этом случае 28 полос анализа одинаковой ширины. Далее их ширина возрастает с ростом частоты так, что они включают соответственно по 3, 6, 12, и, наконец, по 24 коэффициента МДКП каждая. Всего в нашем случае будет 256 коэффициентов МДКП, а полос анализа

50, что соответствует полосе частот входного звукового сигнала 0…23900 Гц.

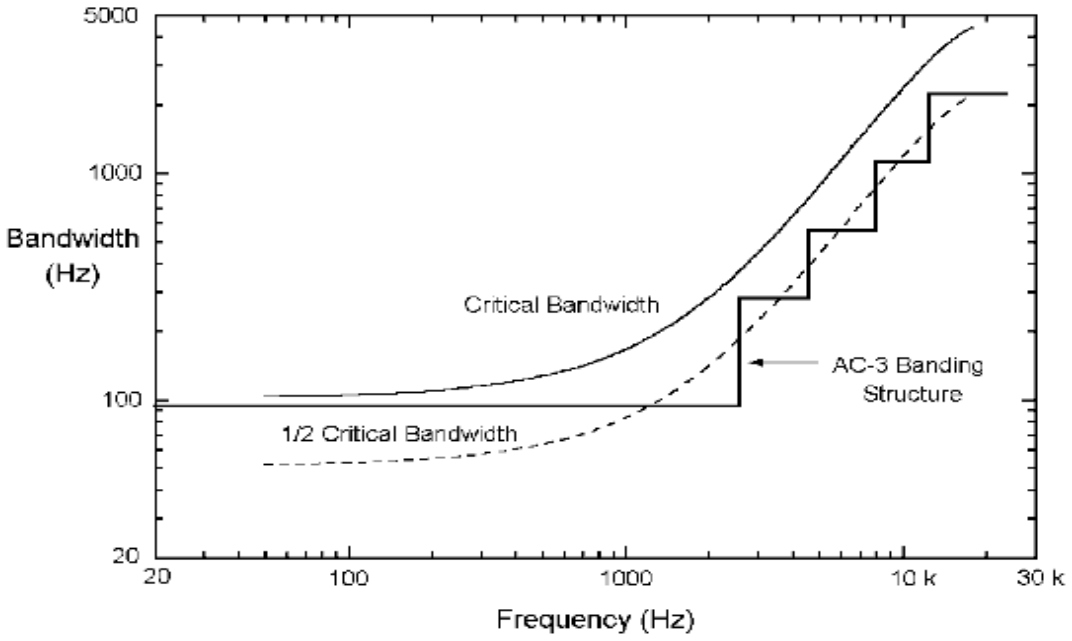


Рис. 1.96. Изменение ширины полос психоакустического анализа в функции от средней частоты для системы *Dolby AC*-3

Изложенное выше поясняет верхний график на рис.1.97. Здесь изображены массивы значений PSD[k] порядков коэффициентов МДКП в полосах психоакустического анализа (синий и зеленый цвета).

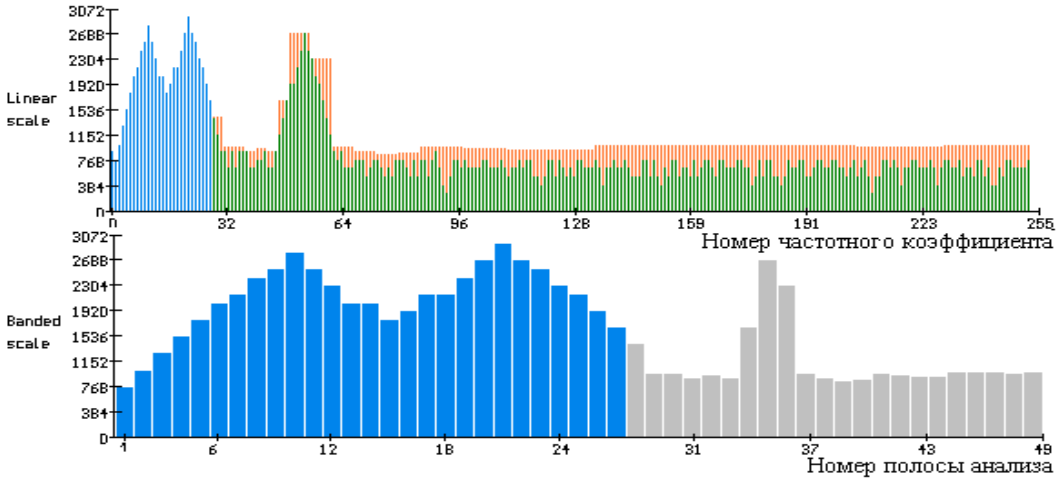


Рис. 1.97.Структура полос психоакустического анализа в системе *Dolby AC*-3

2.Расчет энергии звукового сигнала в полосах психоакустического анализа. Суммарное значение энергии звукового сигнала в каждой полосе

психоакустического анализа в стандарте *Dolby AC*-3 вычисляется по формуле (рис.1.94, блок 2):

$$\log(a+b)=\max[\log(a),\log(b)]+\log[1+\exp(d)],$$

где d = |log(a)-log(b)| является адресом таблицы, значения которой вычислены как: log[1+exp(d)]; log(a) , log(b) – значения порядков соседних (в пределах одной полосы психоакустического анализа) коэффициентов МДКП. В стандарте использован следующий механизм вычислений: сначала берутся значения порядков первых двух коэффициентов МДКП в данной полосе анализа и определяется максимальное значение. К нему добавляется величина равная разности значений порядков этих двух коэффициентов. Величина последней задается в форме табличной функции экспоненциального вида. Затем полученное таким образом суммарное значение порядка записывается в аккумулятор (накопитель) и далее осуществляется сравнение этого вычисленного значения со значением порядка следующего по номеру индекса коэффициента МДКП и т. д. Процесс вычислений повторяется до тех пор, пока не будут использованы все коэффициенты МДКП в данной полосе анализа. На рис.1.97 (нижний график) показан массив значений энергий порядков коэффициентов МДКП в полосах анализа, а на верхнем графике красным цветом тот же самый массив, но вместо номера полосы анализа по оси абсцисс отложены номера коэффициентов МДКП, что позволяет увидеть число коэффициентов МДКП в каждой из полос психоакустического анализа.

3.Выбор прототипа индивидуальной кривой маскировки. Как и в стандартах *MPEG* здесь учитывается маскировка только в частотной области. В основе выбора прототипа кривой маскировки лежат экспериментальные данные, полученные для тонов и заимствованные из работ *E.Zwicker*, *R.Ehmer*, *L.Fielder* и *E.Benjamin*. В качестве примера на рис.1.104 изображено семейство кривых маскировки для тона частотой 3200 Гц. Параметром каждой кривой является абсолютный акустический уровень маскирующего тона, дБ, вычисленный относительно звукового давления $p_0 = 2\cdot10^{-5}$ Па, и равный соответственно 40, 60, 80 и 100 дБ.

Для каждого уровня маскирующего тона (из экспериментальных данных подобных представленным на рис.1.98) определялся относительный порог слышимости шума с полосой частот равной полосе психоакустического анализа. Средняя частота полосы маскируемого шума изменялась. Иными словами, рассматривалась маскировка вне критической полосы слуха, при которой тон маскирует шум с полосой частот примерно равной 0,5 барка. Затем каждая полученная таким путем зависимость нормировалась к уровню маскирующего тона по формуле

$$N_{\mathrm{RNT}}(F)=N_{\mathrm{NT}}(F)-N_{\mathrm{MT}}(F),\ \text{дБ},$$

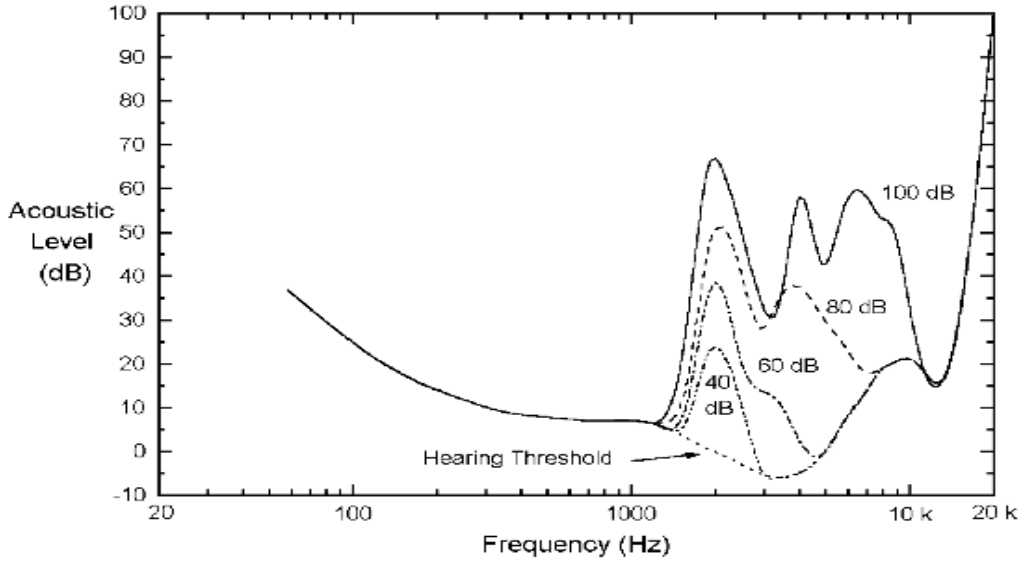


Рис. 1.98. Изменение порога слышимости измерительного тона, маскируемого тоном частотой 3200 Гц.

где $N_{RNT}(F)$ – нормированное по отношению к уровню маскирующего тона значение порога слышимости шума, $N_{NT}(F)$ – порог слышимости шума, маскируемого тоном, $N_{MT}(F)$ – уровень маскирующего тона. Эти вычисления выполнялись отдельно для каждого значения частоты и уровня маскирующего тона.

В качестве примера на рис.1.99 показано семейство таких зависимостей для тона частотой 2000 Гц (тонкие линии). Параметром каждой из

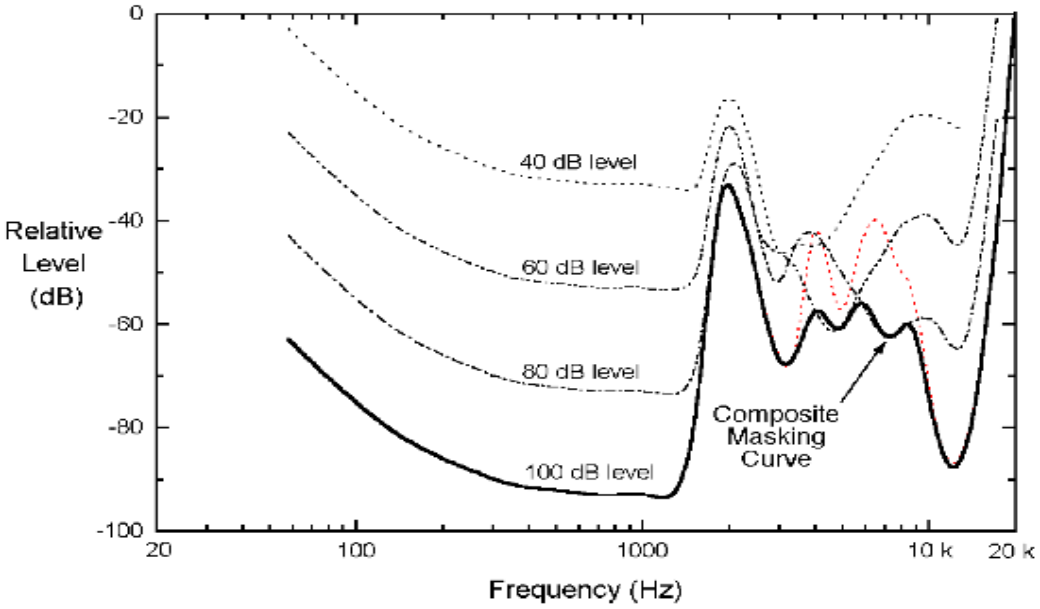


Рис. 1.99. К построению композитной кривой порога слышимости шума, маскируемого тоном частотой 2000 Гц

представленных здесь кривых является уровень маскирующего тона $N_{МТ}$, в данном случае частотой 2000 Гц. Жирной линией здесь показана так называемая композитная кривая маскировки, представляющая собой возможные наименьшие значения для всего представленного здесь семейства нормированных кривых маскировки. В результате этих вычислений разработчиками было в общей сложности получено десять таких композитных зависимостей. Далее эта совокупность композитных кривых маскировки была преобразована в так называемые обобщенные кривые маскировки. Последние представлены на одном графике в виде, показанном на рис. 1.100 (тонкие кривые). По оси абсцисс здесь отложены уже значения высот тона (а не частоты, как это было ранее), деления следуют через 0,5 барка, что соответствует расстоянию между центрами соседних полос психоакустического анализа системы *Dolby AC*-3, а нуль этой новой шкалы соответствует высоте маскирующего тона, выраженной в барках. Параметром каждой такой обобщенной кривой маскировки является частота маскирующего тона. Представленные здесь кривые и есть ничто иное, как развертывающие функции или индивидуальные кривые маскировки.

4.Кодирование параметров обобщенных кривых маскировки. Аппроксимация обобщенных кривых маскировки достаточно сложна. Поэтому далее при их использовании введены упрощения. В системе *Dolby AC*-3 маскировка в сторону низких частот не учитывается. Маскировка в сторону верхних частот (рис.1.100) с точностью вполне достаточной для практики

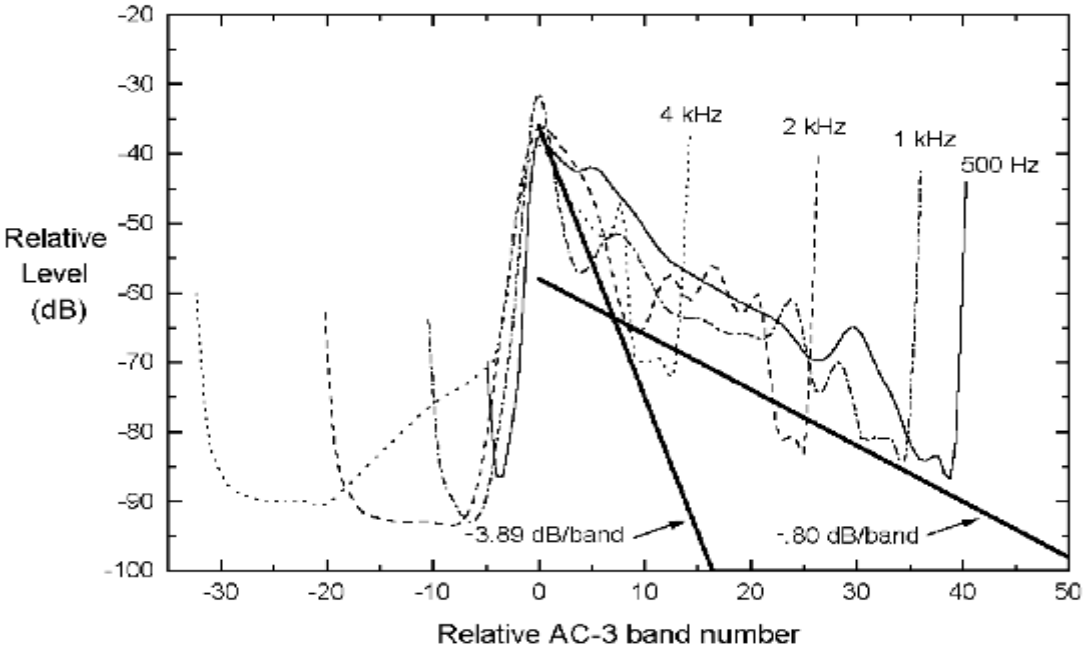


Рис. 1.100. К построению обобщенных кривых маскировки для области частот 500…4000 Гц.

может быть аппроксимирована двумя отрезками прямых линий (жирные линии): *Fast Upwards Masking*-быстро затухающая прямая и *Slow Upwards Masking*-медленно затухающая прямая (рис.1.100 и 1.101). Для их задания в

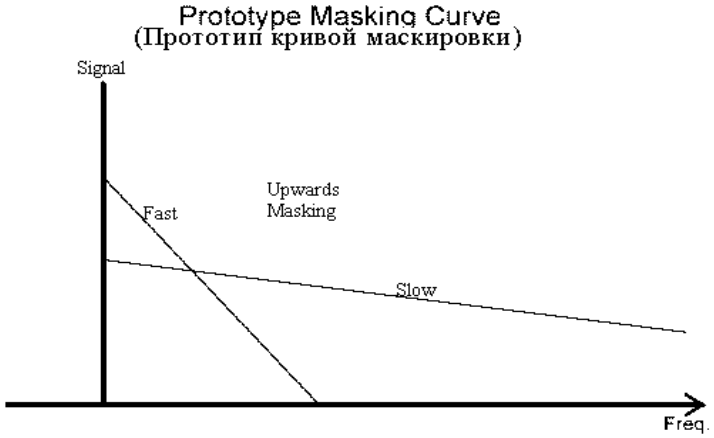


Рис. 1.101. Прототип обобщенной кривой маскировки в системе *Dolby AC*-3

системе *Dolby AC*-3 используются четыре параметра, которые включаются кодером в поле данных психоакустической модели для передачи декодеру и обозначаются как (рис.1.94, блок 3, пункт 2):

Slow Decay – крутизна медленно затухающего сегмента, диапазон изменений данного параметра составляет от - 0,70 до - 0,98 дБ/полосу психоакустического анализа;

Slow Gain – вертикальное смещение вниз от уровня маскирующей компоненты сигнала для медленно затухающего сегмента, диапазон изменений этой величины составляет от - 49 до - 63 дБ;

Fast Decay – крутизна быстро затухающего сегмента, диапазон изменений крутизны наклона от - 2,95 до - 5,77 дБ/полосу анализа;

Fast Gain – вертикальное смещение быстро затухающего сегмента прямой от максимального уровня маскирующей компоненты, диапазон изменений этой величины равен от - 6 до - 48 дБ.

5.Синтез обобщенной кривой маскировки. Отрезки аппроксимирующих прямых линий в кодере Dolby AC-3 синтезируются с помощью двух рекурсивных фильтров, включенных параллельно. Результирующее значение глобального порога маскировки $E_M(к)$ в полосе психоакустического анализа *к* определяется как наибольшее значение из выходных отсчетов этих двух фильтров. Математически, процедуру вычисления значений глобального порога маскировки можно записать следующим образом (рис.1.94, блок 4, пункт 3):

$$x_0(k) = [x_0(k) - d_0(k)] \oplus [E_C(k) - g_0(k)],$$
$$x_1(k) = [x_1(k) - d_0(k)] \oplus [E_C(k) - g_1(k)],$$
$$E_M(k) = max\ (x_0, x_1),$$

где $E_C(k)$ – энергия звукового сигнала в полосе психоакустического анализа k; $d_0(k)$ и $d_1(k)$ – крутизна наклона, соответственно для быстро (*Fast Decay)* и медленно (*Slow Decay)* затухающего сегментов обобщенной кривой маскировки; $x_0(k)$ и $x_1(k)$ – вертикальное смещение сегментов от максимального уровня *к*-той спектральной компоненты сигнала, соответственно для быстро и медленно затухающего сегментов (*Fast Gain[ch] и Slow Gain);* $\oplus$ – оператор "*log-addition*" (логарифмическое сложение); в алгоритме *Dolby AC*-3 этот оператор заменен на оператор *max***.**

Далее полученные значения $E_M(k)$ корректируются с целью учета влияния уровня маскирующего сигнала на величину порога маскировки. После коррекции значения глобального порога маскировки в каждой полосе психоакустического анализа сравниваются с величиной абсолютного порога слышимости и выбирается наибольшее из этих двух значений. В результате выполнения этих операций получается *результирующая кривая маскировки,* определяющая допустимые значения мощности шумов квантования в каждой из полос психоакустического анализа. Минимально допустимое для каждой полосы психоакустического анализа отношение сигнал-шум квантования SNR_n, дБ, вычисляется как разность уровней энергий полезного сигнала и шумов квантования, лежащих на пороге слышимости. При расчете энергии полезного сигнала используются только значения порядков коэффициентов МДКП. Значение $SNR_n[k]$, приведенное к одному коэффиценту МДКП данной полосы психоакустического анализа вычисляется как $SNR_n[k] = SNR_n/k$, где k – число коэффициентов МДКП в n-ой полосе психоакустического анализа. Именно этот массив данных, образует *кривую глобального полога маскировки*, которая и определяет число бит, выделяемых на кодирование мантисс коэффициентов МДКП. Пример данной кривой показан на рис.1.102, верх, кривая синего цвета. Коэффициенты

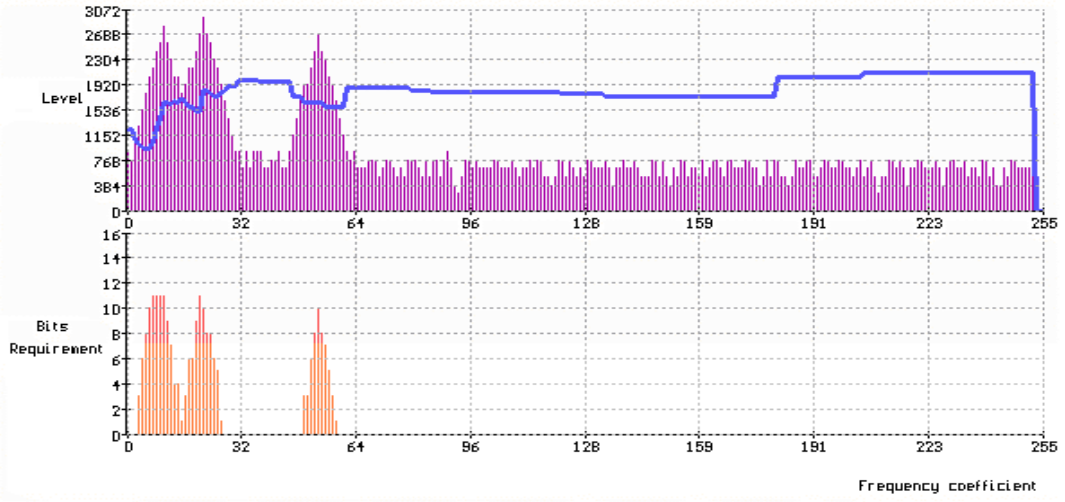


Рис. 1.102. К процедуре выделения бит при кодировании мантисс коэффициентов МДКП

МДКП, расположенные ниже этой кривой, не кодируются и не передаются на приемную сторону системы передачи, ибо лежат ниже относительного порога слышимости. Ни нижней части рис.1.102 приведено число бит, которое требуется выделить для кодирования мантисс каждого из тех коэффициентов МДКП, передача которых необходима на приемную сторону цифровой системы передачи.

При вертикальном смещении кривой глобального порога маскировки изменяется отношение сигнал-шум (*SNR*), а следовательно, и число выделяемых бит на кодирование мантисс коэффициентов МДКП. Возможны «грубое» и «плавное» смещение кривой глобального порога маскировки. Достигается это изменением параметров *CSNR* и *FSNR[ch]*. Причем "плавное" смещение кривой с величиной шага 3/16 дБ достигается изменением параметра *FSNR[ch]*, а "грубое" смещение кривой с шагом 3дБ – изменением параметра *CSNR*. Суммарная величина смещения кривой глобального порога маскировки *(SNROFFSET)* относительно ее первоначального исходного положения определяется формулой:

$$SNROFFSET = ((CSNR - 15)\cdot 16 + FSNR[ch])\cdot 4.$$

Здесь параметры *SNROFFSET, CSNR* и *FSNR[ch]* выражены в целочисленных значениях шкалы, о которой было сказано выше. Значение параметра *FSNR[ch]* определено стандартом *Dolby AC*-3 отдельно для сигнала каждого кодируемого канального сигнала [*ch*], в то время как значение параметра *CSNR* одинаково для всех кодируемых сигналов.

6.Процедура выделения бит. Перед началом итерационного процесса выделения бит кривая глобального порога маскировки устанавливается в верхнее максимальное положение (рис.1.103). При этом оба параметра

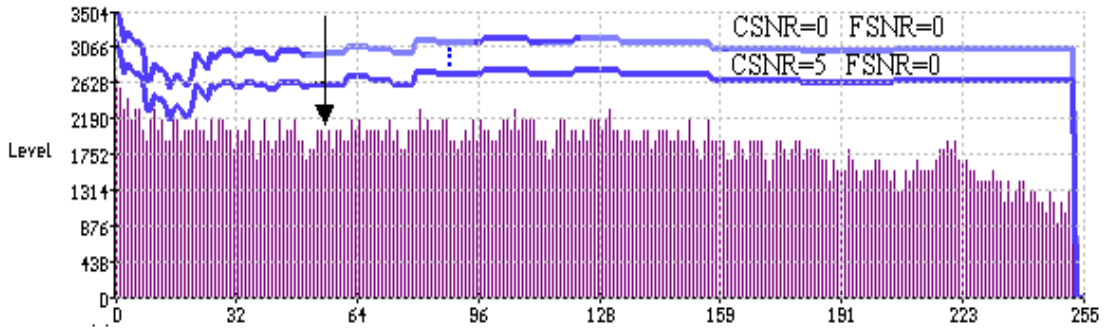


Рис. 1.103.Смещение кривой глобального порога маскировки с шагом 3 дБ (первый цикл итерационного процесса)

FSNR[ch] и *CSNR* принимают нулевые значения. Далее кривая маскировки смещается вниз на 3 дБ при каждом шаге итерации, что соответствует "грубому" ее смещению. При этом значение параметра *CSNR* увеличивается на 1 при каждом шаге итерации. Процесс смещения кривой вниз повторяется до тех пор, пока число выделяемых на кодирование мантисс коэффициентов МДКП бит не превысит доступного их числа, определяемого установленной скоростью цифрового потока. В этом и состоит первый цикл итерационного процесса.

Второй цикл итерационного процесса заключается в медленном смещении кривой глобального порога маскировки, но уже вверх от последнего ее положения. При этом величина шага смещения составляет уже существенно меньшее значение равное 3/16 дБ, что соответствует изменению параметра *FSNR[ch]* на 1 при каждом таком шаге итерации (рис.1.104). Процесс "плавного" смещения этой кривой вверх повторяется до тех пор, пока число бит, выделяемых для кодирования мантисс, станет меньше или равно их доступному количеству для установленной скорости передачи.

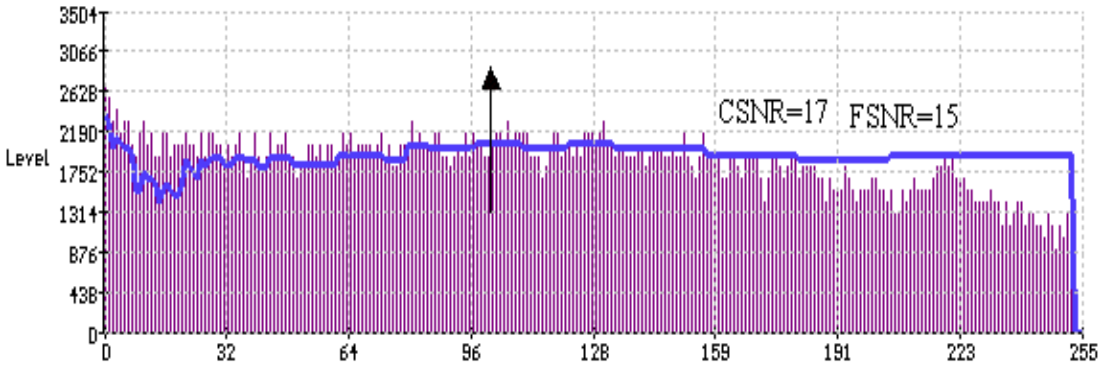


Рис. 1.104. Смещение кривой глобального порога маскировки с шагом 3/16 дБ (второй цикл итерационного процесса)

Помимо вышеперечисленных в процессе выделения бит используется дополнительно еще один параметр, обозначенный как *Floor*. Значением этого параметра устанавливается определенный уровень смещения кривой глобального порога маскировки, ниже которого она опуститься не может, что при определенных ситуациях может вызвать спрямление расчетной кривой глобального порога маскировки (рис. 1.105).

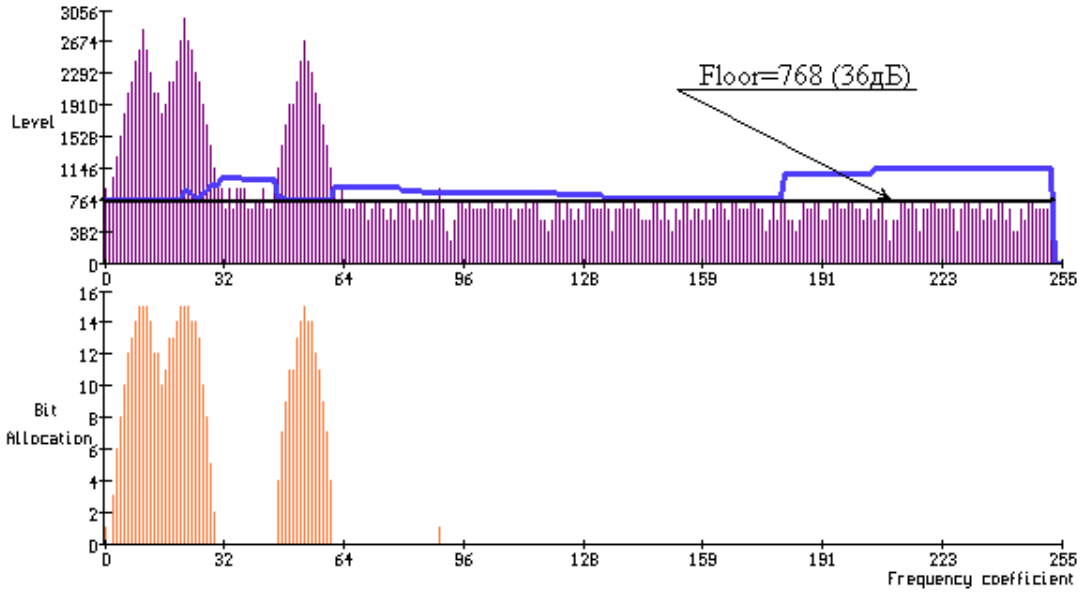


Рис. 1.105.К пояснению влияния *Floor*-параметра

2. КОМПРЕССИЯ ЦИФРОВЫХ АУДИОДАННЫХ

2.1. Краткая характеристика стандартов *MPEG*

Предварительно дадим краткие дополнительные пояснения к стандартам группы *MPEG* (*Moving Pictures Experts Group*).

Стандарт MPEG-1 ISO/IEC 11172-3 рекомендуется для кодирования высококачественных моно- и двухканальных стереофонических сигналов. Он предусматривает использование трех значений частот дискретизации ЗС равных 32, 44,1 и 48 кГц.

Стандарт MPEG-2 ISO/IEC 13818-3 – это обратно совместимая с *MPEG*-1 версия метода кодирования ЗС различных форматов: 1/0, 2/0, 2/1, 3/1, 3/2, 5.1, звуковых сигналов матричных систем фирмы *Dolby Lab* (*Dolby-Stereo*, *Dolby-Surround* и *Dolby-Pro-Logic* и т.п.). Он использует (дополнительно к уже имеющимся значениям в *MPEG*-1) частоты дискретизации равные 16, 22,05 и 24 кГц. Сами же основные алгоритмы компрессии здесь такие же, как и в *MPEG*-1.Более простыми методами кодируются здесь сигналы так называемого многоканального расширения.

Стандарт MPEG-2 ISO/IEC 13818-7 AAC предназначен для высококачественного (*indistinguishable quality*) в соответствии с требованиями *EBU* кодирования ЗС в полной полосе частот (до 20 кГц) при скоростях передачи около 64 кбит/с.

Стандарт MPEG-4 ISO/IEC 14496-3 ориентирован на мультимедиа-приложения. Он спроектирован так, чтобы расширить возможности между мультимедиа терминалами мобильного доступа низкой сложности до высококачественных звуковых систем. Он использует базовые идеи и алгоритмы кодирования, уже определенные в стандарте *MPEG*-2 *ISO/IEC* 13818-7 *AAC*, а также новые идеи, основанные на параметрическом представлении музыкальных и речевых сигналов.

В стандартах *MPEG* предусмотрено несколько уровней (слоев) компрессии цифровых данных: Layer 1, Layer 2 и Layer 3.

Layer 1 (слой 1) - рекомендуется для применения в профессиональной области, в системах записи-перезаписи с высоким студийным качеством с достаточной емкостью памяти. Он характеризуется небольшой сложностью и невысокой степенью редукции аудиоданных. Основные параметры: скорость цифрового потока 192...256 кбит/с, коэффициент компрессии около 4-х, задержка сигнала при обработке около 20 мс.

Layer 2 (слой 2) - потребительская область применения, высококачественное радиовещание; ему соответствует средняя сложность и средняя степень компрессии цифровых аудиоданных. Основные параметры: рекомендуемая скорость цифрового потока 128 кбит/с при кодировании ЗС с

полосой частот равной 40...15 кГц; коэффициент компрессии 6; задержка сигнала при обработке 40...50 мс.

Layer 3 (слой 3) -рекомендуется для передачи ЗС по сети *ISDN* в профессиональной области со средним качеством, *Internet*-вещания, отличается высокой сложностью и характеризуется следующими параметрами: скорость цифрового потока 64 кбит/с при полосе звукового сигнала 40...15 кГц, время задержки при его обработке более 50 мс. Программное обеспечение кодеров MP3 распространяется бесплатно. С повышением скорости цифрового потока для улучшения качества звучания некоторые подпрограммы перцепционного кодирования в *MP*3 не используются. В результате уже при скорости 128 кбит/с (стерео) качество звука такое же, как и у *CD*, тогда как при этом в среднем на один отсчет выборки при его кодировании приходиться всего 1,17 бита. В настоящее время в сети *Internet* распространяются файлы MP3 записанные со скоростью 192 Кбит/с (стерео). В этом последнем случае формат *MP3* уже можно отнести к компрессированию практически без потерь.

Начиная с 1994 и по апрель 1997 года, была проведена работа в рамках MPEG-2 по созданию стандарта, определяющего алгоритм сжатия сигналов многоканальной стереофонии, не отвечающего требованию обратной совместимости. Необходимость создания такого алгоритма обусловлена тем, что требование обратной совместимости предусматривает использование процедур матрицирования (кодер) и дематрицирования (декодер), которые, как показывают исследования, являются источниками дополнительных искажений, ухудшающих качество, при кодировании звукового сигнала.

В связи с большим распространением мобильных технологий, интернета, цифрового телевидения и радиовещания уменьшение скорости передачи цифровых аудиоданных при кодировании сигналов обычной стереофонии и особенно многоканальной стереофонии форматов 5.1, 6.1 и т.п. остается попрежнему весьма актуальным. В 2004 году группой MPEG был инициирован ряд работ, связанных с повышением эффективности кодеков с компрессией цифровых аудиоданных. Эти работы были завершены в 2006 году, затем часть из них вошла в стандарт ISO/IEC 23003-1: 2007 Part 1: *MPEG D Surround*. Исследования, проводимые в этом новом направлении, получили общее название *Spatial Audio Coding (SAC)*. Сюда вошли такие алгоритмы как *Joint Stereo Coding* (*M/S Stereo Coding*, *Intensity Stereo Coding*), *Parametric Stereo*, *Binaural* Cue *Coding*, *Spatia*l *Audio Coding* (аудиоформаты 5.1 и выше).

В монографии рассматриваются только алгоритмы стандартов *MPEG*-4 *ISO/IEC* 14496-3 и *MPEG* D *Surrund*, применяемые в системе цифрового радиовещания DRM

2.2.Общие сведения о стандарте *MPEG* -4 *ISO/IEC* 14494-3

Стандарт *MPEG* -4 *ISO/IEC* 14496 разработан группой *MPEG* в 1997 году для радиовещания и приложений, охватывающих мультимедийные системы: от несложных мобильных с упрощенными терминалами оконечных устройств до профессиональных высококачественных. Здесь рассматриваются только алгоритмы и инструменты, относящиеся к обработке аудиоинформации, изложенные в части 3 стандарта *MPEG* – 4 *ISO/IEC* 14496-3.

Алгоритм кодирования звуковых сигналов, изложенный в стандарте MPEG-4 ISO/IEC 14496-3, позволяет получить скорости цифрового потока для натуральной речи и музыки, изменяющиеся в диапазоне от 2 до 64 кбит/с.

Стандарт *MPEG*-4 *ISO/IEC* 14496-3 включает в себя три разных алгоритма сжатия (рис. 2.1).

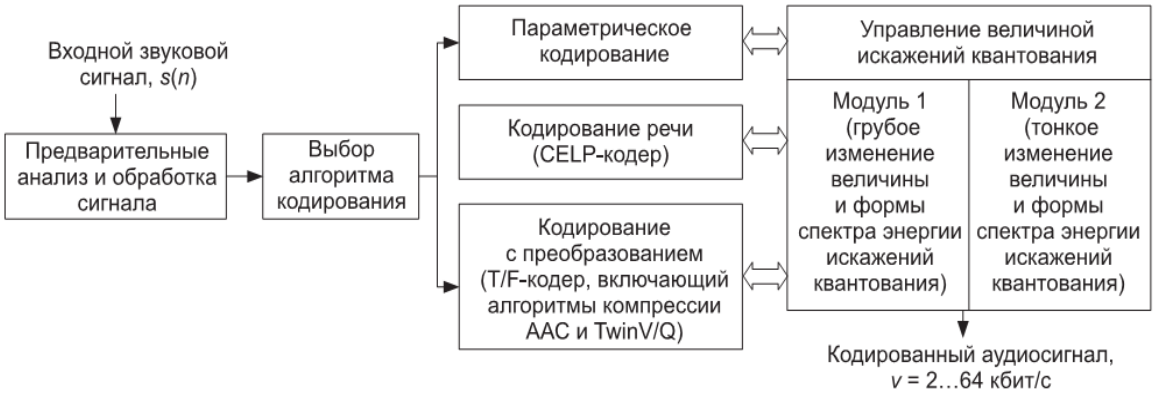


Рис. 2.1. Алгоритмы компрессии цифровых аудиоданных стандарта MPEG - 4 ISO/IEC 14496-3

-параметрическое кодирование (*MPEG* - 4 *ISO/IEC* 14496-3, *Subpart* 2), используется при скоростях цифрового потока, изменяющихся в пределах от 2 до 8…10 кбит/с

-техника *CELP* (*Code Excited Linear Predictive*) кодирования (*MPEG* - 4 *ISO/IEC* 14496-3, *Subpart* 3), используется при кодировании речевых сигналов при скоростях цифрового потока, изменяющихся в пределах от 4 до 24 кбит/с;

-техника *T*/*F* (*Time/Frequency*) кодирования с преобразованием, включающая алгоритм компрессии *AAC* (*Subpart* 4) и *Twin V/Q* кодирование (часть *w1903twq*); техника *T*/*F* используется для кодирования высококачественных звуковых сигналов при скорости цифрового потока, изменяющейся в пределах от 8…10 до 64 кбит/с.

Кроме того, стандарт *MPEG* - 4 *ISO/IEC* 14496-3 включает дополнительно:

-методы синтеза звуковых сигналов на основе *MID*I протокола (Sub-*part* 5);

-синтез речи на основе *TTS* алгоритма (кодер, выполняющий преобразование письменного текста в ясную и четкую речь, *Subpart* 6);

-всевозможные инструменты (фильтрацию, ограничение, динамическое регулирование уровней, микширование и т.п.), благодаря которым пользователь, манипулируя цифровыми потоками, может создавать разнообразные звуковые эффекты;

-возможность изменения пользователем: скорости передачи цифровых данных, полосы частот звукового сигнала, уровня сложности кодера и декодера, помехоустойчивости в отношении цифровых ошибок;

-поддержку многоязычных текстов, различных алгоритмов синтеза речевых и музыкальных сигналов, ряд других функций, всего того, что представляется важным для мультимедиа приложений.

В информационной части стандарта *MPEG* - 4 приведены две психоакустические модели. Обе они могут быть использованы в любом алгоритме компрессии стандарта *MPEG* - 4. Процеруры обработки ЗС при психоакустическом анализе подробно изложены в разд. 1 данной книги

2.3.Алгоритм кодирования *AAC*

Алгоритм компрессии *AAC* (*Advanced Audio Coding*) базируется на учете опыта, накопленного при разработке алгоритма компрессии Layer 3 стандартов *ISO/IEC* 11172-3 и 13818-3, поддерживает все известные звуковые форматы: моно, обычное стерео, разновидности систем *Dolby*, пятиканальный звуковой формат 5.1.

В отличие от *MPEG* - 2 *ISO/IEC* 13818-3 в алгоритме *AAC* (рис. 2.2) расширен набор возможных частот дискретизации: 8; 11,025; 16; 22, 05; 24; 32; 44,1; 48; 64; 88,2 и 96 кГц; изменены форма и длины оконных функций: здесь используются окна Кайзера-Бесселя вместо синусных: длинное, включающее 2048 отсчетов ЗС и короткое – соответственно 256 отсчетов ЗС, что обеспечивает более высокое разрешение по частоте, при этом в обоих случаях используется 50% перекрытие выборок отсчетов ЗС. Кодированию подвергаются коэффициенты МДКП, однако несколько изменена форма кривой компрессии при неравномерном квантовании, применены иные книги кодов Хаффмана, чем в стандарте 13818-3 Layer 3. Кроме того, здесь имеется возможность программным путем заблокировать от 1 до 3 субполос, т.е. не кодировать коэффициенты МДКП в этих субполосах, изменяя, таким образом, полосу передаваемых частот. Этот режим используется в адаптивной конфигурации, о которой речь пойдет ниже. Как и в Layer 3 адаптивное управление величиной искажений квантования выполняется с помощью двух итерационных циклов: внутреннего и внешнего.

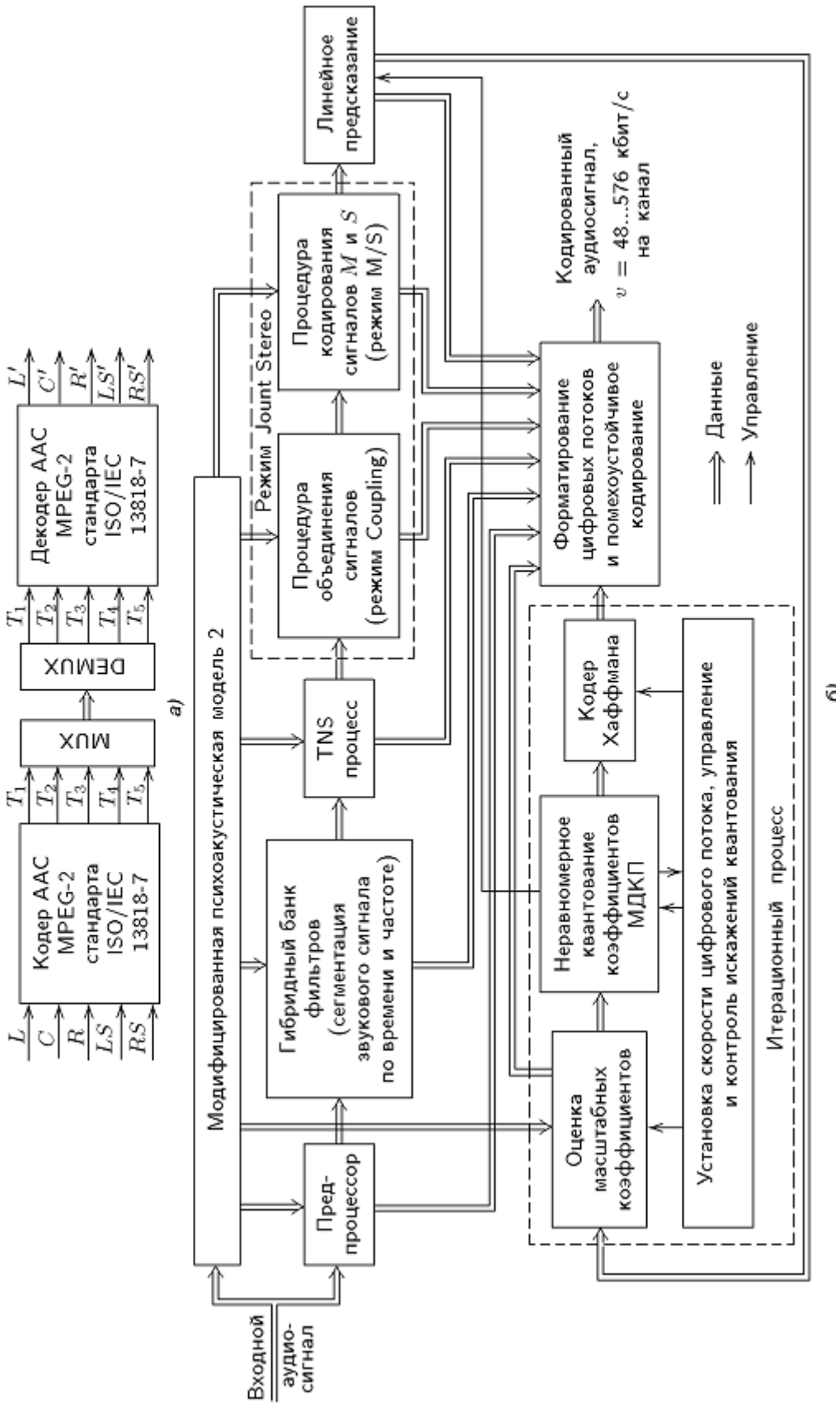


Рис. 2.2. Структурная схема кодера MPEG - 2 AAC (б) стандарта ISO/IEC 13818-7

При кодировании сигналов многоканального расширения используются здесь более простые алгоритмы.

Например, при кодировании сигналов, не имеющих резких выбросов временной функции по амплитуде, весьма эффективным оказывается алгоритм линейного предсказания (*Prediction)*, рис. 2.3. Предположим, что на

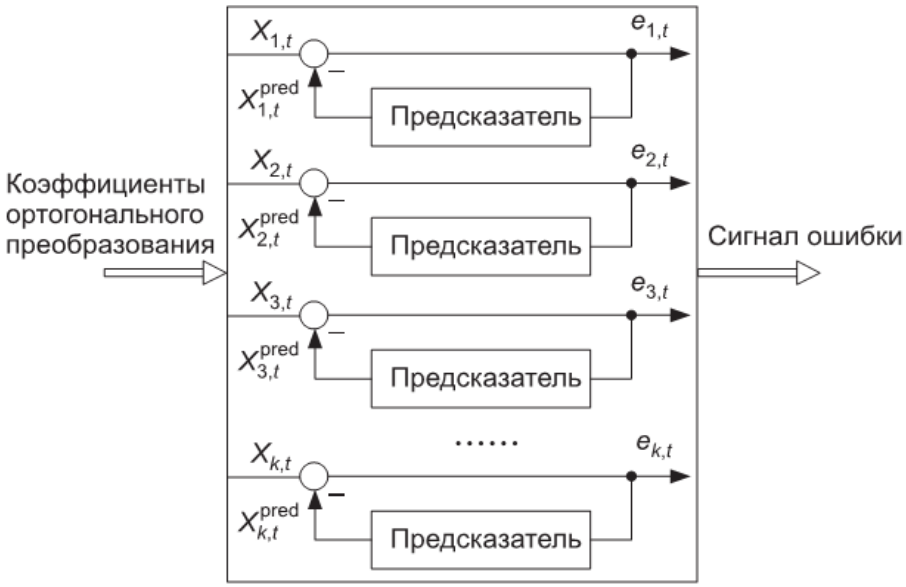


Рис. 2.3. Упрощенная структурная схема блока линейного предсказания

вход блока линейного предсказания второго порядка поступает выборка коэффициентов МДКП. Процедура линейного предсказания предусматривает кодирование не самих квантованных значений коэффициентов МДКП, а так называемого сигнала ошибки

$$e_{k,t} = X_{k,t} - X_{k,t}^{pred} ,$$

где $e_{k,t}$ – значение сигнала ошибки для *k*-ого коэффициента преобразования текущего аудиофрейма *t*; $X_{k,t}$ – значение *k*-ого коэффициента преобразования текущего аудиофрейма *t*; $X_{k,t}^{pred}$ – значение предсказанного значения *k*-ого коэффициента предсказания текущего аудиофрейма *t*; при этом

$$X_{k,t}^{pred} = a_1 X_{k,t-1} + a_2 X_{k,t-2} ,$$

где a_1, a_2 – коэффициента предсказания; $X_{k,t-1}, X_{k,t-2}$ – значения *k*-ого коэффициента преобразования в предыдущих двух аудиофреймах. На коди-

рование сигналов ошибки требуется меньшее число бит, чем на кодирование квантованных значений коэффициентов МДКП.

В алгоритме *AAC* для повышения качества алгоритма компрессии цифровых данных применены специальные процедуры минимизации, точнее говоря управления микроструктурой искажений квантования внутри каждой из субполос (так называемая техника *TNS- Temporal Noise Shaping*). Эта процедура применяется при кодировании отрезков звукового сигнала, имеющих значительные изменения амплитуды сигнала в пределах выборки.

На вход блока *TNS* поступает выборка коэффициентов МДКП, после чего выполняется процедура линейного предсказания, но в отличие от предыдущего случая предсказание выполняется для каждого коэффициента МДКП в рамках текущего аудиофрейма в соответствии с выражениями

$$e_{k,t} = X_{k,t} - X_{k,t}^{pred},$$

$$X_{k,t}^{pred} = \sum_{n=0}^{order} a_n X_{k-n,t},$$

где a_n – коэффициент предсказания; $order$ – порядок предсказания.

С другой стороны блок *TNS,* как это вытекает из названия, формирует временную структуру искажений квантования. При выполнении линейного предсказания огибающая искажений квантования при определенном значении порядка предсказания начинает хорошо повторять форму огибающей кодируемого сигнала.

При линейном предсказании учитывается не только корреляция между отсчетами многоканального сигнала, но и форма спектра шумов квантования и его изменение во времени.

Заметим, что блоки *TNS* и *Prediction* реализованы в стандарте на основе цифровых фильтров. В блоке предварительной обработки сигнала, так же, как и в одноименном блоке *MPEG*-1 *ISO/IEC* 11172-3 Layer 3, используется техника динамического изменения длины преобразования. Критерием изменения длины преобразования является значение психоакустической энтропии, вычисляемое в психоакустической модели.

В алгоритме *AAC* изменены процедуры объединения субполосных сигналов при их кодировании (*Coupling*). В нем предусмотрена, как и в более ранних стандартах группы *MPEG*, возможность работы кодера в режиме *M*/*S* кодирования, когда кодированию в субполосах подвергаются не сигналы L и R стереопары, а их сумма $M = (L+R)/\sqrt{2}$ и разность $S = (L-R)/\sqrt{2}$.

Введены уточнения и дополнительные процедуры при расчете глобального порога маскировки в психоакустической модели кодера *AAC*. Однако и здесь основой является модифицированная психоакустическая модель 2, как и в *Layer* 3.

В зависимости от вычислительной сложности и области применения в стандарте *ISO/IEC* 13838-7 *AAC* три возможных конфигурации.

Основная конфигурация (*Main profile*).Она используется, когда вычислительная сложность алгоритма не является сдерживающим фактором при реализации кодека. При данной конфигурации в кодере не используется банк *PQMF*-фильтров. Вся последовательность из 2048 временных отсчетов ЗС непосредственно подается на блок ортогонального преобразования с 50-ти % перекрытием. Порядок предсказания блока *TNS* составляет 20.

Конфигурация пониженной сложности *(Low Complexity profile).* Здесь также не используется банк *PQMF* - фильтров, и, кроме этого, не используется блок линейного предсказания, порядок предсказания блока *TNS* сокращен с 20 до 12.

Адаптивная конфигурация *(Scalable Sampling Rate).* При данной конфигурации кодера в отличие от двух предыдущих используется банк *PQMF*-фильтров, блок предсказания по-прежнему не используется, а порядок предсказания блока *TNS* составляет 12. Кроме того, не всегда требуется передавать сигнал в полосе частот 20...20000 Гц, а иногда это и невозможно в силу ограниченной пропускной способности канала связи, поэтому стандарт определяет так называемую адаптивную конфигурацию, позволяющую изменять полосу передаваемых частот звукового сигнала. Например, можно передать звуковой сигнал в следующих полосах частот: от 20 до 6000 Гц, от 20 до 12000 Гц, от 20 до 18000 Гц.

Тестовые прослушивания показали, что алгоритм компрессии *AAC* обеспечивает так называемое прозрачное кодирование при скорости цифрового потока 64 кбит/с на канал. При звуковом формате 5.1 искажения, вызванные компрессией, лежат ниже порогов их слуховой заметности уже при суммарной скорости цифрового потока 320...384 кбит/с.

2.4.Параметрическое кодирование звуковых сигналов в стандарте *MPEG* - 4

Идея параметрического кодирования представлена на рис. 2.4. Исходный ЗС выборки *s(n)* в блоке сепарации (цикл анализа-синтеза) в соответствии с базовой моделью звукового сигнала разделяется на тональные и шумоподобные составляющие. После этого тональные компоненты подразделяются на *гармонические* (находящиеся в кратном соотношении с частотой основного тона) и *индивидуальные высокого уровня*, где это условие не выполняется. Далее оцениваются значения текущих частот, фаз и амплитуд каждой из тональных компонент, а для шумоподобных составляющих рассчитываются их уровни энергии в определенных полосах частот. Дополнительно могут быть определены параметры амплитудных огибаю-

Рис. 2.4. Идея параметрического кодирования звукового сигнала

щих и условия продолжения выделенных тональных компонент из текущего аудиофрейма в следующий. Значения перечисленных выше параметров квантуются и кодируются минимально-возможным количеством бит. Требуемое для их кодирования число бит определяется с помощью психоакустического анализа. В декодере по значениям переданных параметров синтезируется исходный звуковой сигнал.

Обобщенная структурная схема параметрического кодера стандарта *MPEG*-4 *ISO/IEC* 14496-3 представлена на рис. 2.5. В блоке предварительного анализа и сепарации входной звуковой сигнал разделяется на две компоненты (части): речевую и музыкальную. Для кодирования каждой из них

Рис. 2.5. Обобщенная структурная схема параметрического кодера стандарта *MPEG* - 4 *ISO/IEC* 14494-3, часть *w1903par*

используется свой алгоритм, реализуемый соответственно в кодерах *HVXC* (*Harmonic Vector Excitation* – возбуждение вектора гармоник) и *HILN* (*Harmonic and Individual Lines plus Noise* – гармонические и индивидуальные тональные составляющие плюс шумоподобные компоненты). Это разделе-

ние может быть выполнено вручную или автоматически. В настоящее время поддерживается автоматическое переключение между речевыми и музыкальными компонентами (частями) сигналами, позволяя использовать *HVXC* - кодер только для кодирования речи, а *HILN* - кодер – только для кодирования музыки. В обоих случаях каждый кодер (*HVXC* и *HILN*) содержит два основных блока, один из которых служит для выделения и оценки параметров сигнала, а другой – для квантования и кодирования их значений с учетом свойств слуха.

Формирователь цифрового потока позволяет работать как в режиме чередования (или только в режиме *HVXC*, или только в режиме *HILN*), так и в комбинированном (смешанном) режиме, когда возможно переключение кодеров при переходе от одного сегмента звукового сигнала к другому.

Заметим, что до недавнего времени параметрическое представление использовалось только при кодировании речевых сигналов, более простых по своей структуре, чем музыкальный сигнал. Однако в последние годы благодаря успехам вычислительной техники, математического моделирования, психофизики и электроники параметрическое представление все чаще начинает применяться и при кодировании высококачественных звуковых сигналов, обеспечивая большую степень компрессии цифровых данных.

Параметрическое кодирование музыкальных сигналов, обладая весьма сложными процедурами обработки, и требующее при реализации существенно больших вычислительных затрат, позволяет получить скорость цифрового потока 16...24 кбит/с при достаточно хорошем качестве.

Алгоритм кодирования *HILN*

Основной принцип *HILN* - кодера состоит в анализе входного сигнала с целью извлечения (выделения) описывающих этот сигнал составляющих. Параметры этих составляющих оцениваются по величине, затем квантуются, кодируются и передаются (записываются) в виде потока цифровых данных. В декодере на основе выделенных и переданных кодером параметров генерируются эти составляющие, необходимые для синтеза выходного сигнала.

Кодер HILN содержит два основных блока. При кодировании исходный звуковой сигнал делится на последовательные сегменты (выборки). Для каждого такого сегмента выделяется и затем кодируется набор параметров, возможно более полно описывающих звуковой сигнал в этом сегменте. Благодаря такому параметрическому описанию возможен широкий диапазон изменения скоростей передачи, частот дискретизации и длин самих сегментов. Обычно используется длина сегмента (выборки) равная 32 мс.

Напомним, что при передаче звукового сигнала по телефонной паре частота дискретизации обычно равна 8 кГц. Скорость цифрового потока данных в этом случае при параметрическом кодировании может быть уменьшена до 6 кбит/с. Для широкополосных сигналов, например, музыкальных, частота дискретизации составляет чаще всего 48 кГц, при этом параметрическое кодирование позволяет уменьшить скорость цифрового потока до 24 кбит/с.

Структурная схема HILN-кодера. В принципе для описания ЗС могут использоваться различные наборы параметров и разные способы для их выделения, разные методы его синтеза. С этой точки зрения входной сигнал при его сепарации должен быть разделен на составляющие вполне определенным образом, известным декодеру, то есть с использованием вполне конкретной модели сигнала. Эта процедура выполняется при анализе сегмента (выборки) сигнала (рис. 2.6). В процессе анализа должны быть

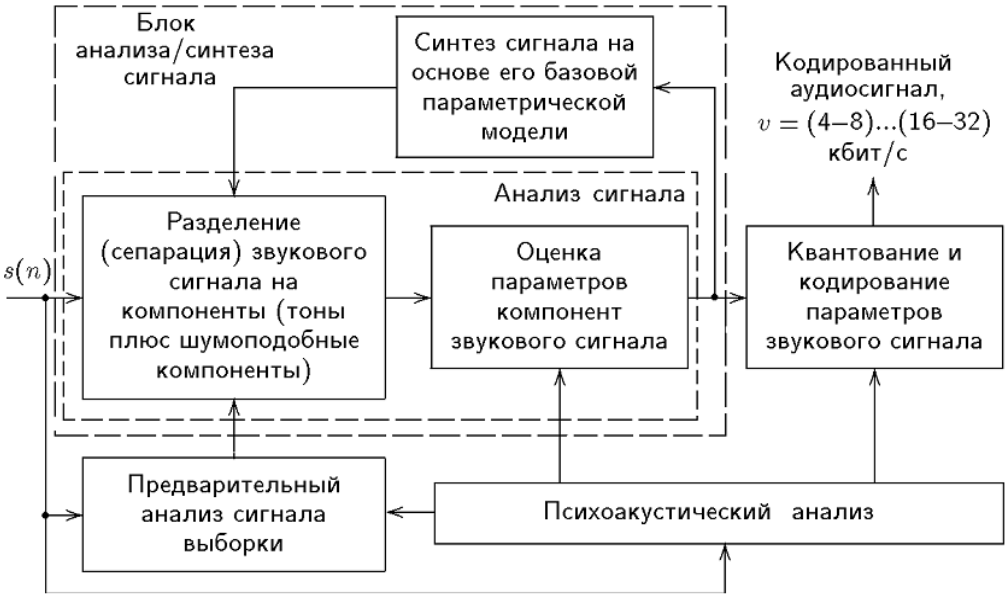


Рис. 2.6. Упрощенная структурная схема кодера, реализующего метод параметрического кодирования звукового сигнала (стандарт *MPEG* - 4 *ISO/IEC* 14496-3)

выделены составляющие сигнал компоненты, и после этого оценены их параметры в соответствии с *базовой параметрической моделью* звукового сигнала, взятой за основу при реализации такого кодера. Блоки выделения и оценки параметров исходного сегмента звукового сигнала (выделены желтым цветом на рис. 2.6) рассматриваются здесь как этапы анализа, позволяющего выделить, а затем и оценить значения каждого из параметров сегмента исходного сигнала при его параметрическом описании.

Точность сепарации входного сигнала может быть повышена при помощи *«петли анализа/синтеза»*. Находящийся в этой цепи блок синтеза реконструирует кодируемый сегмент сигнала, используя для его синтеза на-

бор значений выделенных параметров. Далее оба сигнала исходный и сгенерированный (реконструированный) поступают на вычитающее устройство, где рассчитывается сигнал ошибки. После этого значения выделенных параметров *уточняются* с целью минимизации сигнала ошибки. Блоки разделения и оценки параметров исходного сигнала дополнительно получают также данные от блока *предварительного анализа*, что позволяет сделать оценку параметров сигнала более точной.

После выделения и оценки значения параметров квантуются и кодируются. Оба этих процесса выполняются с учетом результатов психоакустического анализа выделенного сегмента сигнала.

Психоакустическая модель кодера обрабатывает входной сигнал для получения информации о значимости выделенных параметров с точки зрения слухового восприятия. Иначе говоря, не все выделенные параметры кодируются и передаются декодеру, а только тех из тональных и шумоподобных компонент сигнала, которые лежат выше порога слышимости, то есть оказывают влияние на слуховую оценку сигнала. Кроме этого психоакустическая модель используется и для анализа синтезируемого сигнала, позволяя получить информацию необходимую, для работы блока «*Оценка параметров компонент звукового сигнала*», рис. 2.6.

В блоке анализа HILN-кодера выборка звукового сигнала делится на три составляющие:

-*основной тон и кратные ему гармонические составляющие,*

-*индивидуальные тональные составляющие не кратные основному тону* и

-*шумовые компоненты.*

Для каждой из этих составляющих оцениваются их параметры:

-для тональных составляющих: *частота, амплитуда и фаза*;

-для шумовых компонент: *форма спектра и уровень энергии в субполосах кодирования.*

Итак, периодические компоненты сигнала после их выделения разделяются на гармонические и негармонические составляющие. Последние в стандарте названы как индивидуальные составляющие. Дополнительно могут быть определены такие параметры как изменение формы огибающей спектра шумовых компонент и изменение уровней тональных и шумоподобных компонент при переходе от текущего сегмента сигнала к последующему.

Разделение сигнала на составляющие и оценка их параметров выполняются в параметрическом кодере следующим образом. Сначала оценивается частота основного тона текущего сегмента сигнала. Затем оцениваются параметры значимых тональных составляющих. После выделения эти составляющие классифицируются как *гармонические составляющие* и (или) *индивидуальные составляющие,* в зависимости от значения их частоты по отношению к частоте основного тона. После выделения всех тональ-

ных составляющих и исключения их из спектра исходного сигнала оставшаяся его часть рассматривается как шумоподобный сигнал. При этом форма его спектра и энергия в субполосах кодирования описываются соответствующим набором параметров.

В стандарте *MPEG* - 4 предусмотрено также использование и так называемого *интегрированного параметрического кодера*, что подразумевает одновременное использование средств *HVXC* и *HILN* при кодировании одного и того же сегмента сигнала. Поясним сказанное. Если входной сигнал является, например, смесью речевого сигнала и музыкального фона, то *HILN* - кодер может быть использован для выделения и кодирования только значимых индивидуальных тональных составляющих, то есть не находящихся в кратном соотношении с частотой основного тона. Оставшаяся часть сигнала, состоящая из гармонических и шумоподобных компонент, кодируется только средствами *HVXC* - кодера.

Теперь рассмотрим все эти процедуры подробнее.

Предварительный анализ сигнала. Для улучшения моделирования переходных процессов, часто имеющих место во входном звуковом сигнале, перед циклом *анализа/синтеза* (*Analysis/Synthesis Loop*), рис. 2.6, выполняется процедура предварительного анализа.

Обычно блок предварительного анализа решает две задачи. Прежде всего, он для каждого сегмента входного сигнала определяет, длинное или короткое окно анализа следует использовать при анализе/синтезе сигнала. Решение о длине окна принимается, исходя из расчета соотношения значений максимальных амплитуд сигнала, как в текущем сегменте, так и в интервале, выходящем за его границы и охватывающем половину длин предшествующего и последующего сегментов. При этом если отношение максимальных амплитуд вне текущего сегмента и внутри его превышает заданный порог, то используется короткая оконная функция. В противном случае применяется длинная оконная функция. В качестве длинного окна используется функция Ханна, вдвое превышающая длину аудиофрейма. Короткое окно – это прямоугольно-подобное окно с гладкими переходами в его начале и в конце. В результате сглаживания переходов, короткое окно оказывается немного длиннее самого сегмента выборки.

Кроме выбора длины оконной функции в блоке предварительного анализа вычисляется огибающая сигнала выборки и рассчитывается набор описывающих ее параметров. Эти данные используются затем в блоке анализа/синтеза.

Структурная схема формирователя огибающей сигнала выборки изображена на рис. 2.7. При вычислении огибающей входной сигнал предварительно пропускается через полосовой фильтр ПФ (*band pass filter*), который позволяет ослабить составляющие ЗС на очень высоких и очень низких частотах. После фильтрации сигнал выборки подвергается преобразо-

ванию Гильберта (*Hilbert transformer*) для получения мнимой составляющей исходного сигнала. Здесь все его спектральные компоненты получают

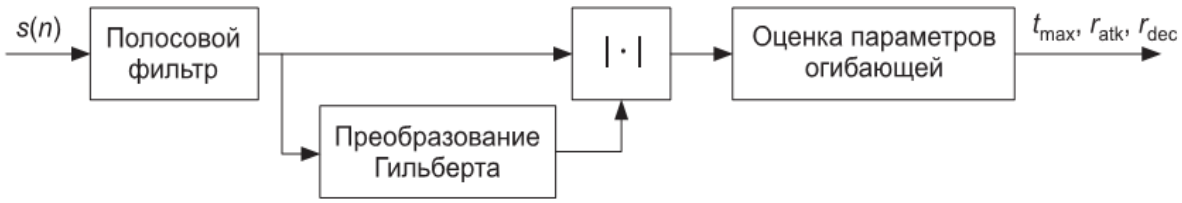


Рис. 2.7. Структурная схема формирователя огибающей сигнала выборки

сдвиг по фазе относительно исходного сигнала на 90^0. В следующем блоке |·| рассчитывается огибающая сигнала выборки. Она нормируется относительно ее максимального значения в пределах текущей выборки.

Обычно ПФ и преобразование Гильберта реализованы в виде так называемых КИХ-фильтров. Частотные характеристики этих фильтров для частоты дискретизации 8 кГц показаны на рис. 2.8. Заметим, что полосовой фильтр необходим для подавления сигнала на тех частотах, где преобразование Гильберта не может быть правильно вычислено.

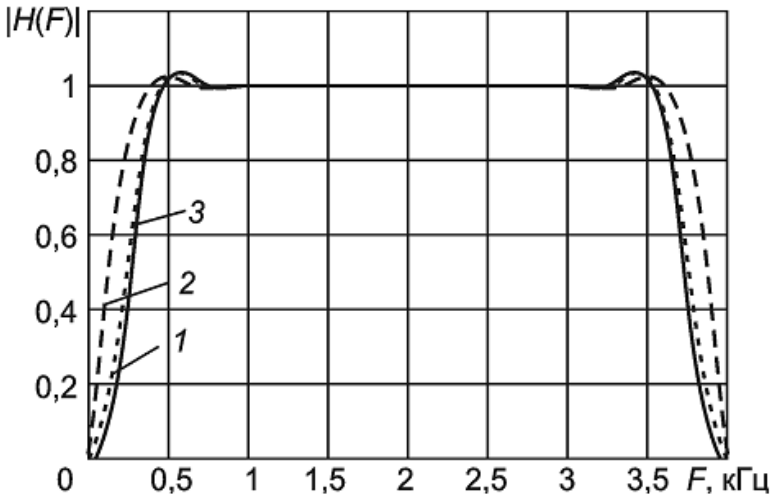


Рис. 2.8. Частотные характеристики полосового фильтра (точечная линия), преобразователя Гильберта (пунктирная линия) и системы в целом, включающей оба устройства (сплошная линия)

Модель огибающей ЗС, используемая в кодеке, включает фазу атаки и фазу затухания сигнала (соответственно до и после выброса). Для описания формы огибающей используются три параметра. Их значения вычисляются в следующем блоке данного устройства (блок *attack and decay parameter estimation;* оценка параметров огибающей на рис. 2.7):

$t_{\max}$ – временная позиция (положение на оси текущего времени) максимума амплитуды, определяющая конец фазы атаки и начало фазы затухания;

r_{atk} – скорость атаки, определяемая углом наклона кривой атаки;

r_{dec} – скорость спада, определяемая углом наклона кривой затухания.

Значение t_{max} оценивается относительно (в пределах) длины выборки. Значения параметров, определяющих атаку и спад, с целью упрощения процедуры кодирования, задаются углами наклона соответствующих им отрезков прямых линий, а также значением постоянной амплитуды огибающей до и после атаки. Если наклон кривой затухания (значение угла) достигает нуля до окончания выборки, то значение огибающей для оставшейся части сегмента устанавливается равным нулю. Аналогичным образом задается нулевое значение огибающей до момента начала атаки. В качестве примера на рис. 2.9 показаны исходная и сгенерированная по вычисленным параметрам $t_{\max}$, r_{atk} и r_{dec} огибающие для одной выборки сигнала.

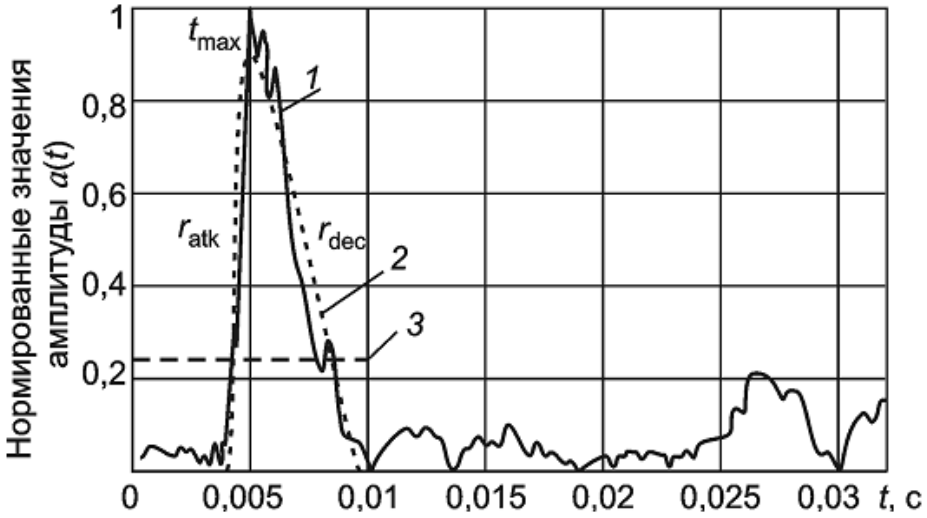


Рис. 2.9. Огибающая сигнала текущего аудиофрейма: оригинал (сплошная линия), ее сгенерированные значения по оцененным параметрам (точечная линия)

При определении параметров огибающей сначала находят величину $t_{\max}$. Параметр $t_{\max}$ соответствует точке, в которой огибающая сигнала достигает своего максимального значения впервые. Далее вычисляется среднеквадратичное значение амплитуд огибающей. Для определения значений r_{atk} и r_{dec} используется линейная аппроксимация огибающей сигнала выборки. Заметим, что скорость атаки r_{atk} – это угол наклона линии, проходящей через максимум огибающей в момент времени $t_{\max}$, которая

наилучшим образом аппроксимирует временную функцию огибающей в области атаки до момента достижения ею максимума. Для повышения точности аппроксимации используется весовая функция. Ее значение равно единице для тех участков, на которых значения огибающей меньше среднеквадратичного значения, на остальных участках значение весовой функции непрерывно возрастает с увеличением амплитуды самой огибающей. Величина весовой функции увеличивается также и по мере приближения текущего времени к значению $t_{\max}$. Скорость спада r_{dec} – это угол наклона линии, наилучшим образом аппроксимирующей огибающую выборки звукового сигнала после момента времени $t_{\max}$. В данном случае используется та же весовая функция, что и при оценке атаки. Однако эта линия не обязательно должна проходить через впадину после максимума $t_{\max}$. Заметим, что только значение угла наклона аппроксимирующей функции используется в качестве параметра r_{dec}; при этом ее вертикальное положение линии не учитывается.

Если скорость атаки и скорость спада огибающей остаются в пределах выборки ниже порогового значения, (это означает, что в пределах текущего фрейма не имеется никаких быстрых изменений амплитуды), то рассчитывается угол наклона такой линии, которая наилучшим образом аппроксимирует изменение огибающей сигнала на протяжении всей выборки. В данном случае весовая функция аппроксимирующего выражения зависит только от амплитуды исходной огибающей. Если градиент изменения амплитудных значений огибающей сигнала больше нуля, то вычисляется только параметр r_{atk}, при этом значение $t_{\max}$ соответствует концу выборки. Таким образом моделируется медленно увеличивающаяся в течение выборки амплитуда огибающей ЗС. Соответственно, если значение градиента ее изменения меньше нуля, то вычисляется параметр r_{dec}, а значение $t_{\max}$ соответствует началу фрейма.

По этим трем параметрам ($r_{atk}, r_{dec}, t_{\max}$) генерируется огибающая звукового сигнала текущей выборки.

Оценка частоты основного тона. Прежде всего, оценивается частота основного тона текущего сегмента звукового сигнала. Для грубой первоначальной ее оценки вможно использовать, например, технику *кепстрального анализа*.

Для получения *кепстра* сначала входной сигнал взвешивается окном Хемминга длины, равной удвоенной длине сегмента и центрированного относительно текущего сегмента. Для взвешенного таким образом сигнала вычисляется спектр, далее берется его модуль, затем полученный амплитудный спектр сигнала выборки логарифмируется и взвешивается окном

$$w(F) = (1+\cos(2\pi F/f_Д))/2, \qquad \text{где} \quad 0 <= F <= f_Д/2.$$

и после этого выполняется обратное преобразование Фурье.

Выполнение этих процедур на практике встречает целый ряд вычислительных сложностей, связанных с эффектом наложения частот. Однако сегодня они преодолены. Заметим, что здесь мы говорили о действительном кепстре, но, как и в случае частотного спектра, кепстр может быть и комплексным. В этом случае после логарифмирования комплексного спектра сигнала выборки мы получим два слагаемых. При этом действительная часть этой суммы представляет собой логарифм спектра амплитуд, а мнимая – характеризует фазовый спектр. Но в данном случае нас интересует действительный кепстр $C(k)$, когда обратное преобразование Фурье выполняется только над спектром амплитуд:

$$C(k) = F^{-1}\{w(k) \cdot \ln |X(k)|\} = F^{-1}\{ w(k) \cdot \ln [|F\{s(n)|]\},$$

где $F\{\cdot\}$ и $F^{-1}\{\cdot\}$ – символы прямого и обратного дискретного преобразования Фурье; $X(к)$ – спектральная компонента сигнала выборки с индексом $к$; $s(n)$ – отсчет выборки сигнала с номером n, $w(k)$ – оконная функция.

В качестве примера на рис. 2.10,*а* показан сегмент (выборка) вокализованной речи, взвешенный окном Хемминга, на рис. 2.10,*б* представлен

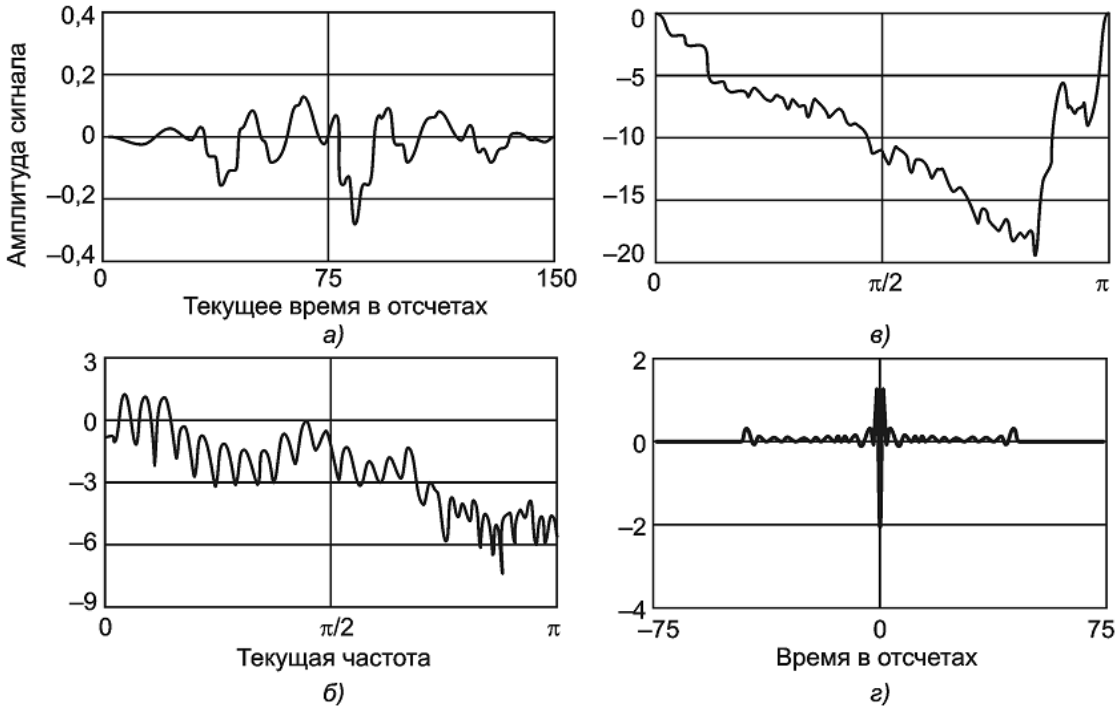


Рис. 2.10. К оценке частоты основного тона выборки звукового сигнала: *а* — взвешенный окном Хемминга фрагмент звукового сигнала; *б* — логарифм модуля кратковременного преобразования Фурье;
в — значения текущей фазы для данного отрывка; *г* — кепстр

логарифм модуля дискретного преобразования Фурье для этого взвешенного сегмента сигнала, на рис. 2.10,*в*- значения фазы, имеющие разрывы, но здесь прерывистый характер фазы устранен специальной процедурой. На рис. 2.10,*г* мы видим кепстр данного сегмента речевого сигнала. Если в исходном сигнале существуют периодические (тональные) компоненты (гармонические и индивидуальные), то вычисленный кепстр будет иметь локальные пики. Наибольший по величине локальный максимум соответствует основному тону. Местоположение пика дает хорошую оценку частоты основного тона. Однако, основная частота, определенная посредством этой техники, является лишь начальной (грубой) оценкой частоты основного тона.

Оценка параметров тональных составляющих. Для этой цели используется *цикл анализа/синтеза*. С помощью данной процедуры рассчитываются параметры тональных компонент. Расчет производится итерационно.

В качестве первого шага (рис. 2.11) выполняется грубая оценка параметров тональных компонент выборки. Если посмотреть на спектр сигнала

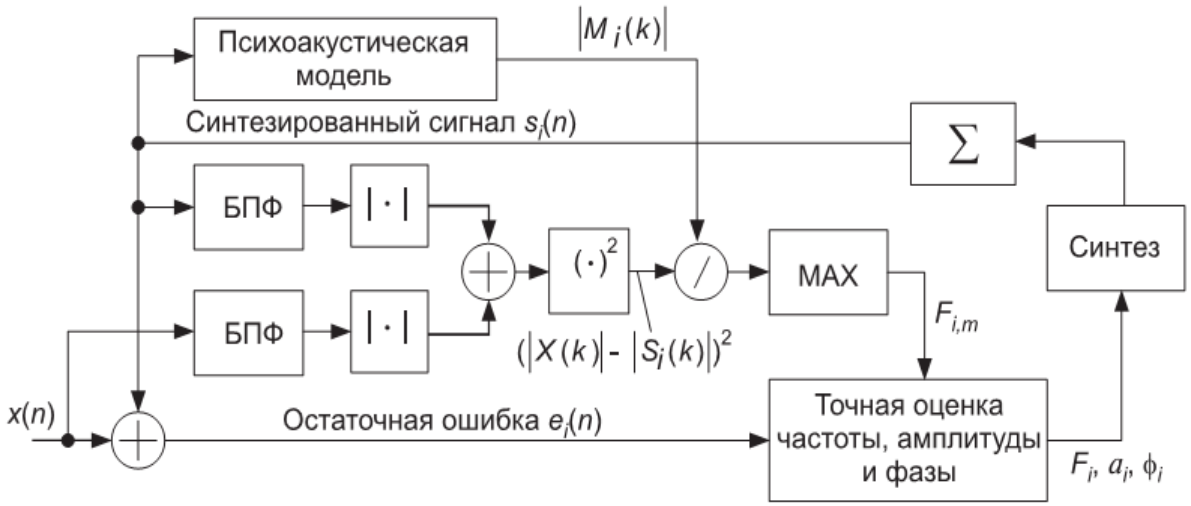


Рис. 2.11. Цикл *анализа/синтеза* на основе метода синтеза «отдельной тональной составляющей»

ошибки $(|X(k)| - |S_i(k)|)^2$, то в местах расположения тональных компонент, отличающихся существенно по уровню от соседних, будут иметь место максимумы АЧХ. А это и есть области частот, где расположены основной тон, гармонические компоненты и тональные компоненты высокого уровня не кратные по частоте основому тону. Это первая ступень грубой оценки их параметров. Далее выполняется их высокоточная оценка. Это выполняется с помощью схемы, изображенной на рис. 2.12. О существе этой процедуры и о самом рисунке будет сказано чуть ниже, ибо представленный

здесь алгоритм является общим при оценке частоты всех тональных компонент. Затем на основе точных значениях частот гармонических составляющих вычисляется точная оценка частоты основного тона $F_{от}$ и так

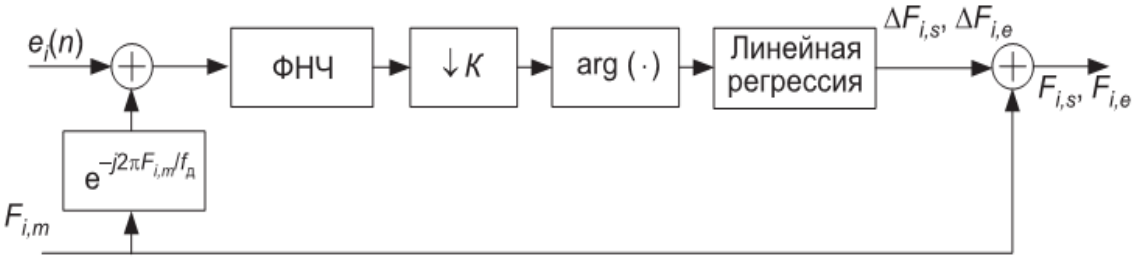


Рис. 2.12. Структурная схема устройства для точной оценки частотыосновного тона сигнала выборки

называемое *расширение по частоте* $\Delta F_г(i+1)$, минимизирующее значение ошибки между реальными частотами гармонических составляющих и вычисленными их значениями в соответствии с выражением

$$F_г(i) = F_{от}(i+1)\cdot(1+ \Delta F_г(i+1), \text{ где } i=0,1,2,\ldots(n-1), \qquad (2.1)$$

где n – общее число гармонических составляющих в спектре сигнала текущей выборки. Оно определяется шириной полосы частот сигнала ΔF и основной частотой $F_{от}$ сигнала текущей выборки $n = floor(\Delta F/ F_{от})$. Здесь запись $floor(\Delta F/F_{от})$ означает, что берется целая часть частного от деления. Для гармонических составляющих устанавливается соответствующий флаг.

Флаг гармонической составляющей для каждой из них устанавливается, если использование огибающей при оценке их амплитуд приводит к меньшей остаточной ошибке, по сравнению с тем, когда огибающая сигнала для этой цели не используется. Если относительное изменение частоты основного тона между предыдущим и текущим сегментами не превышает 15%, то устанавливается флаг продолжения данной составляющей.

Во втором шаге из входного сигнала посредством цикла *анализа/синтеза* выделяются значимые тональные составляющие. Для оценки значимости (с позиций слухового восприятия) каждой из этих компонент используется психоакустическая модель, что позволяет расположить их в порядке убывания значимости. Если частота выделенной тональной компоненты ближе к частоте гармонической составляющей, вычисленной из (2.1), то эта выделенная составляющая классифицируется как гармоническая. В противном случае она классифицируется как индивидуальная составляющая. Цикл *анализа/синтеза* прерывается, если было выделено требуемое число индивидуальных составляющих или если оставшиеся компоненты сигнала не могут быть правильно смоделированы посредством тональных составляющих. Отношение числа выделенных гармонических

компонент к общему числу выделенных тональных составляющих передается в кодер в качестве меры «*значимости*» гармонических компонент.

Если в результате выполнения процедуры *анализа/синтеза* менее трех выделенных тональных компонент были классифицированы как гармонические, то они добавляются к перечню индивидуальных составляющих и величине *n* присваивается значение 0. Все гармонические составляющие, которые не были выделены в цикле *анализа/синтеза*, также удаляются из остаточного сигнала. Затем этот остаточный сигнал передается в блок оценки параметров шума.

Оценка параметров тональных компонент. Применяемая в данном случае методика оценки параметров тональных компонент с высоким разрешением по частоте используется отдельно для каждой спектральной компоненты (для каждого коэффициента быстрого преобразования Фурье, БПФ). Наиболее значимая тональная компонента выбирается на каждом шаге итерации с помощью психоакустической модели (ПМ). ПМ позволяет вычислить порог маскировки для синтезируемого (реконструируемого) сигнала, который содержит только те спектральные компоненты, которые были ранее извлечены (найдены и описаны с помощью параметров) из текущей выборки на предыдущих итерационных циклах.

Процедура оценки параметров для каждой *i*-ой тональной компоненты (один проход цикла *анализа/синтеза*) включает следующие операции:

-вычисление разности между амплитудными спектрами БПФ входного $|X(k)|$ и синтезируемого $|S_i(k)|$ сигналов;

-поиск наиболее значимого коэффициента БПФ (наиболее значимой тональной компоненты сигнала), центральная частота которого обозначается $F_{i,m}$;

-дополнительный анализ спектра сигнала выборки вблизи частоты $F_{i,m}$ с большим разрешением, что необходимо для более точной оценки ее частоты;

-оценку амплитуды и фазы выделенной тональной компоненты, а также вычисление параметров огибающей звукового сигнала выборки.

Выбор наиболее значимого коэффициента БПФ (спектральной компоненты), который будет обработан на данной итерации цикла, осуществляется путем вычисления разности между амплитудными спектрами входного и синтезируемого сигналов и поиска максимального отношения квадрата этой разности. При этом, конечно, учитывается и значение порога маскировки $|M_i(k)|$. Он вычисляется для синтезированного сигнала, сгенерированного на основе параметров тональных компонент, уже оцененных на предыдущих итерациях.

При точной оценке частоты используют схему, изображенную на рис. 2.12. Представленный с ее помощью алгоритм позволяет получить более точное значение частоты по сравнению с тем, что дает расчет БПФ, где

мы, как известно, получаем дискретный спектр. Более ранняя методика оценки частоты, основанная на линейной аппроксимации значений текущей фазы, была позже заменена аппроксимацией второго порядка, которая позволяет осуществлять оценку значений частоты выделяемой компоненты, линейно изменяющейся в пределах выборки. Реализуется данная процедура следующим образом. Сигнал ошибки (*residual error,* $e_i(n)$, рис. 2.11), представляющий собой разность между исходным и сгенерированным сигналами, сдвигается по частоте вниз на значение $-F_{i,m}$ путем умножения (x), так как это выполняется в любом преобразователе частоты (рис. 2.12):

$$v(n) = e_i(n) \cdot e^{-i \cdot 2\pi n \cdot F_{i,m}/f_{Д}} \Rightarrow V(F) = E_i\left(F - F_{i,m}\right) \qquad (2.2)$$

(где $f_Д$ - частота дискретизации) так, чтобы ее значение стало равным нулю. Полученный после этого преобразования комплексный сигнал пропускается через фильтр нижних частот (ФНЧ) и затем подвергается процедуре понижения частоты дискретизации $f_Д/K$. Далее оценивается изменение значений фазы $\Delta\varphi(k)$ для полученного комплексного сигнала. Чтобы получить значения фаз, которые могут выходить за интервал [-180^0…+180^0], сначала для каждой пары двух соседних отсчетов комплексного сигнала рассчитывается сдвиг фазы по формуле:

$$\Delta\varphi(k) = \arg\left(w(k) \quad w^*(k-1)\right) \quad \text{äëÿ} \quad k \geq 1$$

Эти значения считаются достоверными, если пониженное значение частоты дискретизации в два раза больше граничной частоты среза ФНЧ.

Полученные для каждого отсчета сигнала значения фаз суммируются

$$\varphi(k) = \sum_{i-1}^{k} \Delta\varphi(i)$$

Полученную в результате этих вычислений зависимость изменения фазы данной компоненты сигнала в пределах выборки аппроксимируют. При этом используется аппроксимация второго порядка, рис. 2.13,*а*. В результате дифференцирования данной кривой получаем зависимость (точечная кривая на рис. 2.13,*а*) изменения частоты анализируемой тональной компоненты (точки на рис. 2.13,*б*), используя выражения:

$$\Delta \hat{F}(t) = \frac{1}{2\pi} \cdot \frac{d}{dt} \hat{\varphi}(t) \qquad \hat{F}(t) = F_{i,m} + \Delta \hat{F}(t) \qquad (2.3)$$

Изменения частоты $\Delta F_{i,e}$, $\Delta F_{i,s}$ и значения частоты $F_{i,e}$, $F_{i,s}$ вычисляются путем подстановки в (2.3) значений времени t_s, t_e, соответствующих началу и концу выборки:

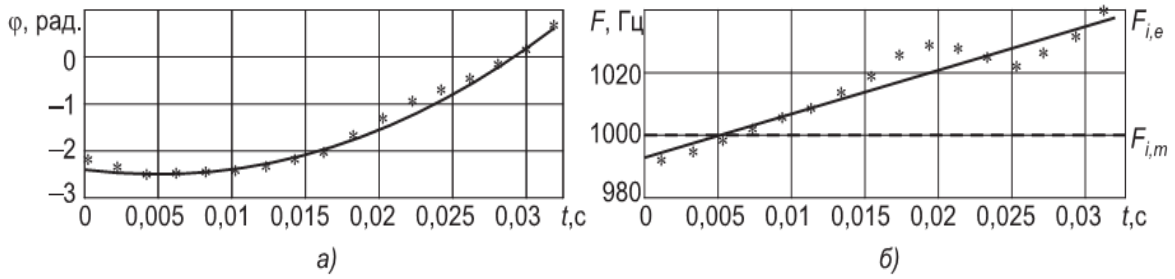


Рис. 2.13. Примеры зависимостей изменения фазы (*а*) и частоты (*б*), полученные с помощью устройства (рис. 2.12) точной оценки частоты основного тона сигнала выборки:сплошные линии — зависимости регрессии; звездочки — результаты вычислений

$$\Delta F_{i,s} = \Delta \hat{F}\left(t_s\right), \quad F_{i,s} = F_{i,m} + \Delta F_{i,s}$$

$$\Delta F_{i,e} = \Delta \hat{F}\left(t_e\right), \quad F_{i,e} = F_{i,m} + \Delta F_{i,e}$$

Заметим, что сплошные линии, изображенные на рис. 2.13 – результат применения к полученным точечным значениям регрессионного анализа. Чаще всего оценка параметров кривых регрессии выполняется с помощью метода наименьших квадратов.

После того как точные значения частот найдены, вычисляются их амплитуды и фазы. Для этой цели рассчитывается комплексное значение коэффициента корреляции сигнала ошибки (*residual error*, рис. 2.11) и тонального сигнала, частота которого изменяется в пределах выборки от $\Delta F_{i,s}$ до $\Delta F_{i,e}$. Модуль найденного значения коэффициента корреляции соответствует амплитуде a_i, а фаза – параметру фазы φ_i выделенной тональной компоненты. Заметим, что наклон линии регрессии для фазовых величин (рис. 2.13,*а*) полученных комплексных отсчетов определяет частотный сдвиг, который добавляется к значению $F_{i,m}$ с целью получения высокоточного частотного параметра F_i. В данной реализации временной сдвиг функции окна простирается в пределах от – 0,32 до + 0,32 длины сегмента. Шаг временного сдвига равен 0,08 длины сегмента, и, таким образом, для линейной регрессии используются 9 значений данных. Иными словами, значения фазы и частоты для каждой тональной компоненты рассчитываются несколько раз в течение выборки (эти значения показаны на рис. 2.13 звездочками).

Заметим, что если на этапе преданализа была рассчитана и сгенерирована огибающая сигнала выборки, то второй набор параметров $a_{i,env}$, $\varphi_{i,env}$ соответствует амплитуде и фазе коэффициента корреляции сигнала ошибки и представленного в комплексной форме тонального сигнала, дополнительно умноженного на синтезированную огибающую. При этом частота тонального сигнала изменяется в пределах выборки также от $F_{i,s}$ до $F_{i,e}$.

Значения частоты тональной компоненты, соответствующие началу $F_{i,s}$ и концу $F_{i,e}$ выборки, используются в последующих итерациях, то есть при расчете параметров другой тональной компоненты, а также и для синтеза сигнала выборки. Это позволяет минимизировать сигнал ошибки для случая, когда частоты отдельных тональных составляющих изменяются в пределах выборки. Кроме того, полученные параметры изменения частоты тональной компоненты ($F_{i,s}$ и $F_{i,e}$) используются также для принятия решения о том, можно ли данную тональную компоненту считать продолжающейся от одной выборки к другой, то есть имеющейся в нескольких выборках.

При последующих операциях кодируется и квантуется для каждой выборки только среднее арифметическое значение частоты тональной компоненты:

$$F_i = \frac{F_{i,s} + F_{i,e}}{2}.$$

Блок синтеза (рис. 2.11) генерирует синусоидальный (тональный) сигнал согласно параметрам $F_{i,s}$, $F_{i,e}$, a_i и φ_i. Если в блоке преданализа вычисляются параметры огибающей, то синтезируется и второй такой же тональный сигнал, но уже на основе параметров $F_{i,s}$, $F_{i,e}$, $a_{i,env}$, $\varphi_{i,env}$, умноженный на синтезированную огибающую.

В блоке сепарации (рис. 2.6) новый сигнал ошибки вычисляется путем вычитания синтезированного сигнала из исходного. Если используется так же и второй синтезированный сигнал (полученный с учетом огибающей), то вычисляется также и второй сигнал ошибки. Затем из двух этих сигналов ошибки (и соответственно из двух наборов параметров) выбирают тот, который обладает самой низкой дисперсией, и в последующих шагах анализа используют только его.

Оценка параметров шумоподобных составляющих. Для оценки параметров остаточного сигнала, прежде всего, вычисляется его спектр. Перед выполнением этой операции взвешивается окном Хемминга. Затем вычисляется автокорреляционная функция полученного сигнала. И далее

вычисляются *LPC*-параметры шума (*LPC – Linear Predctive Coding* или кодирование с линейным предсказанием) с использованием алгоритма Дарбина. После чего *LPC*-параметры преобразуются в *LAR*-параметры. Здесь для моделирования временной функции остаточного сигнала используется фильтр, характеристики которого изменяются в соответствии с рассчитанными *LPC*-параметрами. Кроме этого вычисляется также энергия шумового сигнала. Вычисляется также и отношение энергии остаточного сигнала к энергии исходного сигнала и передается в кодер как мера «значимости» шумоподобной компоненты звукового сигнала.

Кодирование выделенных параметров сигнала текущей выборки. Выделенные параметры гармонических и индивидуальных составляющих, а также и параметры шумоподобной части сигнала выборки кодируются для получения выходного потока цифровых данных *HILN* - кодера.

Квантование параметров гармонических составляющих. Число бит, предназначенных для кодирования параметров гармонических составляющих, зависит от «значимости» каждой из них. Если эта величина мала, то число кодируемых гармоник может быть меньше, чем число выделенных.

Основная частота сигнала выборки квантуется с использованием 2048-шаговой логарифмической шкалы, имеющей диапазон от 20 Гц до 4 кГц. «*Расширенные*» параметры квантуются 5 битами с применением равномерной шкалы с диапазоном изменения от – 0,001 до + 0,001.

Для описания формы спектра компоненты, содержащей только гармонические составляющие, вычисляется функции автокорреляции этой части сигнала. Далее на основе рекурсивного решения полученных автокорреляционных функций рассчитываются *LAR*-параметры, а затем и *LPC*-коэффициенты фильтра, которые приближенно моделируют форму спектра гармонической части исходного сигнала.

Этот процесс близок к *LPC*-моделированию, используемому, обычно, при оценке параметров шумоподобной компоненты текущей выборки сигнала. Кроме *LAR*-параметров, вычисляется также энергия гармонических составляющих исходного сигнала.

Квантование параметров индивидуальных составляющих. В устройстве квантования и кодирования параметры индивидуальных составляющих обрабатываются в том порядке, в каком они поступают из блока *анализа/синтеза*, т. к. он соответствует их значимости при слуховой оценке. В данном устройстве может генерироваться два потока битов: *основной поток битов*, который позволяет генерировать звуковой сигнал так называемого основного качества, и *поток битов улучшения*, который может быть использован в случаях, когда для каких либо других целей требуется разностный сигнал между входным сигналом и выходом декодера, например, для целей масштабирования (изменения) скорости передачи цифровых данных. Основной поток битов обычно содержит значения частот и амплитуд индивидуальных составляющих. Поток битов *улучшения* содержит зна-

чения фаз и дополнительную информацию для более точного квантования значений частот каждой из индивидуальных составляющих и параметров огибающей этой части сигнала.

При этом для каждого аудиофрейма выборки звукового сигнала в соответствии с установленной скоростью цифрового потока передается определенное число бит служебной информации. Первый бит в каждом аудиофрейме – это бит огибающей, определяющий, используется или нет огибающая при кодировании. Если значение этого бита равно 1 (что свидетельствует об ее использовании), то далее следуют 3 параметра огибающей, а затем – собственно параметры данной составляющей.

Заметим, что слуховая система человека не слишком чувствительна к изменениям фазы. Поэтому информация о частоте и амплитуде кодируется и передается в *основном потоке битов* для получения сигнала базисного (основного, стандартного) качества звучания. Но в этом случае необходимо обеспечить получение декодером информации, которая позволяет ему генерировать сигнал, свободный от разрывов фазы на границах сегментов. Следовательно, первый шаг обработки определяет составляющие, которые продолжаются от одного сегмента к другому. Если составляющая должна быть продолжена из предыдущего в следующий сегмент, то вместо абсолютных значений частоты и амплитуды квантуются и передаются далее только их изменения от одного сегмента к другому. Для этого частотные и амплитудные параметры i-ой составляющей m-го сегмента сравниваются с параметрами k-ой составляющей предыдущего $(m-1)$-го сегмента для всех комбинаций i и k. Комбинация составляющих используется, если относительное изменение частоты

$$q_F(i,k) = \frac{\left|F_i(m) - F_k(m-1)\right|}{F_i(m)}$$

не превышает данного предела $q_{F,max}$ и если отношение амплитуд

$$q_a(i,k) = \begin{cases} a_i(m)/a_k(m-1)\text{, ïðè} & a_i(m) \geq a_k(m-1) \\ a_k(m-1)/a_i(m)\text{, ïðè} & a_i(m) < a_k(m-1) \end{cases}$$

лежит в интервале $1 \ldots q_{a,max}$.

В случае, когда имеется более одной возможности продолжения составляющей предыдущего сегмента, то выбирается та из них в предыдущем сегменте, для которой максимален нижеследующий критерий подобия:

$$Q = \frac{q_{F,\max} - q_F(i,k)}{q_{F,\max}} \cdot \frac{q_{a,\max} - q_a(i,k)}{(q_{a,\max} - 1)q_a(i,k)}.$$

Значения частот индивидуальных составляющих квантуются в шкале Барков, а их амплитуды – в логарифмической шкале. Для каждой составляющей предыдущего сегмента в потоке передается так называемый *бит продления*, показывающий, продолжается ли эта составляющая в текущем сегменте. Для новых составляющих индексы квантованных значений частоты и амплитуды кодируются с помощью специальной процедуры, названной в стандарте как *SubDivisionCode (SDC)*. Для всех составляющих, продолжающихся от предыдущего сегмента, кодируются значения разности амплитуд и частот с использованием энтропийного кодирования.

Для сегмента звукового сигнала длиной 32 мс и скорости цифрового потока равной 6 кбит/с в каждом фрейме обычно кодируются параметры 10…17 спектральных составляющих.

В качестве примера на рис. 2.14 показаны отрезки исходного сигнала,

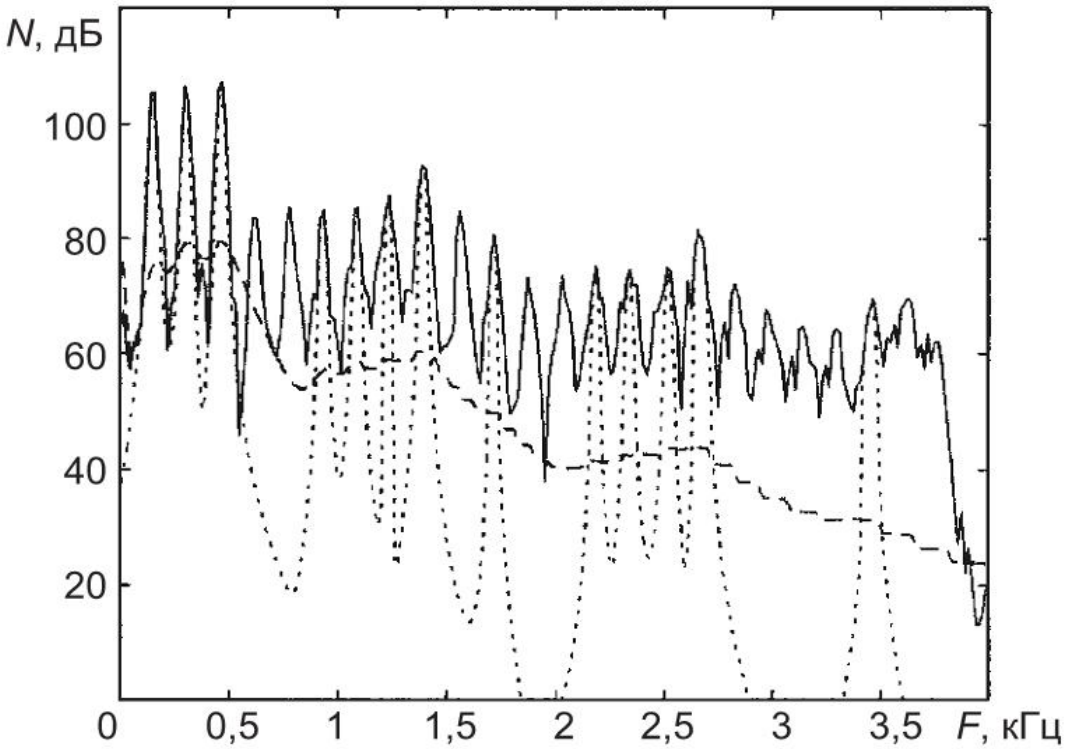


Рис. 2.14. Спектр исходного (сплошная кривая) и синтезированного (пунктирная кривая) сегментов звукового сигнала (мужская речь). Штриховой линией на данном рисунке показан порог слышимости

его синтезированной копии и кривая порога маскировки. Видно, что оригинал и его синтезированная копия имеют достаточно хорошее совпадение в области уровней, превышающих порог слышимости, вычисленный в блоке психоакустического анализа. В отличие от этого на рис. 2.15 представлены

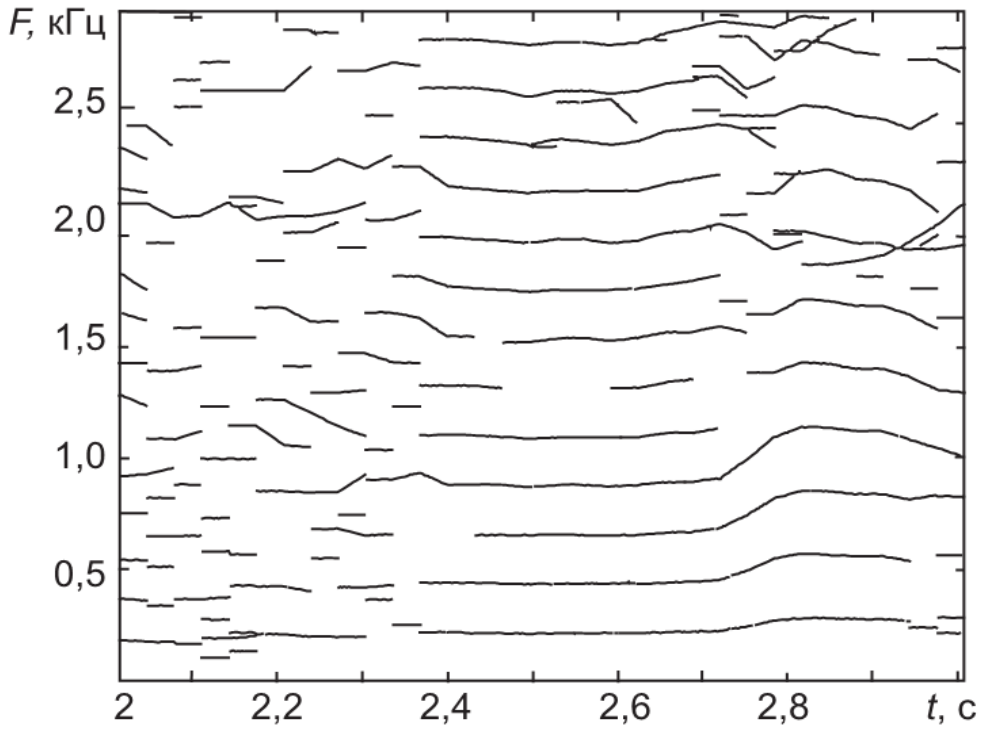


Рис. 2.15. Выделенные в блоке анализа спектральные компоненты звукового сигнала в функции от времени (вокал)

выделенные в блоке анализа тональные компоненты отрезка звукового сигнала. Спектральные компоненты показаны здесь отрезками линий, при этом видно, что большая часть выделенных спектральных компонент продолжается при переходе от одного сегмента звукового сигнала к другому.

Очевидно, что передача абсолютных величин значений амплитуды и фазы требует большего числа бит на составляющую, чем передача их относительных изменений. Кроме того, число составляющих, передаваемых для каждого аудиофрейма, изменяется с целью обеспечения постоянной скорости для *основного потока* битов.

Чтобы были возможны режимы *расширения,* улучшающие качество кодирования, дополнительно генерируется поток *битов улучшения*. Он создается следующим образом:

-если параметры огибающей передаются в *основном потоке битов*, то передаются и дополнительные биты для более точного квантования трех параметров огибающей;

-если составляющая начинается в текущем сегменте, т.е. не является продолжающейся от предыдущего сегмента, и ее частота превышает определенный предел, то передаются дополнительные биты для более точного квантования значения абсолютной частоты;

-для каждой составляющей после квантования передается параметр фазы.

Число битов на сегмент в потоке *битов улучшения* может изменяться, это должно быть принято во внимание при вычислении битов, доступных для кодирования сигнала остаточной ошибки.

Так как положение продолжающейся составляющей в текущем сегменте зависит от положения ее «предшественницы» в предыдущем сегменте, используется алгоритм распределения битов, который удостоверяет, что *N* составляющих, переданных в текущем сегменте, всегда являются теми *N* наиболее вероятными составляющими, которые были выделены блоком *анализа/синтеза*.

Временная задержка в кодере равна 1,5 длинам сегмента. Она складывается из собственно длины сегмента и дополнительной задержки, равной 0,5 длины сегмента, возникающей из-за наложения сдвинутого окна, используемого для оценки частоты.

Квантование параметров шума. Число квантуемых и кодируемых параметров шума зависит от *значимости* (энергии*)* шумоподобной компоненты сигнала. Если она очень низка, то шумовые параметры не передаются. Для более высоких значений меры *значимости* квантуется и кодируется адекватное число *LAR*-параметров. Решение о числе передаваемых *LAR*-параметров может быть принято в кодере, при этом не требуется повторного вычисления этих параметров.

Если установлен флаг *noiseEnvFlag* (данный бит равен 1), то квантуется и кодируется также и дополнительный набор параметров огибающей шума.

Изменение скорости (масштабируемость) потока битов *HILN*-кодера. Благодаря параметрическому представлению звукового сигнала *HILN*-кодер хорошо подходит для задач, при которых требуется масштабирование (изменение) скорости потока битов. При этом скорость потока битов, принимаемого декодером, может быть динамически адаптирована к свойствам канала передачи или может быть выбрана согласно каким-либо другим правилам. В случае, когда требуется передача потока битов с пониженной скоростью, то передаются только параметры наиболее значимых для восприятия компонент сигнала (основной тон, гармонические компоненты, шумопобная часть). В случае же полноскоростного потока битов передаются также и параметры дополнительных компонент сигнала (например, индивидуальных составляющих), которые менее значимы для восприятия, чем передаваемые в низкоскоростном потоке битов. Кроме того, в этом случае передаются также и дополнительные параметры, уточняющие описание параметров сигнала, уже присутствующих в низкоскоростном потоке битов.

Изменение скорости цифрового потока возможно как для основного потока битов, так и для потока расширения, этот режим может использоваться также и при динамически (непрерывно) контролируемом кодировании параметров звукового сигнала.

Изменение скорости потока битов при динамически контролируемом кодировании параметров сигнала. При работе в этом режиме используется тот факт, что процессами выделения и кодирования параметров сигнала в кодере можно управлять независимо. Параметры, сгенерированные устройством их выделения, могут подаваться одновременно на множество устройств кодирования, каждое из которых генерирует поток битов со своей скоростью. Это очень удобно, так как сложность *HILN*-кодера определяется главным образом устройством выделения параметров. Возможно также сохранение в отдельном файле неквантованных параметров, сгенерированных устройством их выделения. В этом случае устройство кодирования параметров может быть использовано для генерации только тех параметров, сохраненных в этом файле, которые обеспечивают получение потока битов с требуемой в текущий момент времени скоростью.

Кодирование смеси речь/музыка

В стандарте *MPEG* - 4 *ISO/IEC* 14496-3 для кодирования звуковых сигналов с очень низкими скоростями передачи, колеблющимися от 2-х до 8-ми кбит/с, используется так называемый интегрированный параметрический кодер, включающий два набора средств, предназначенных для кодирования речевых и неречевых компонент звуковых сигналов соответственно:

-средства *HVXC* - *Harmonic Vector Excitation* («возбуждение вектора гармоник»), предназначенные для кодирования речевых сигналов со скоростями от 2-х до 4-х кбит/с;

-средства *HILN* - *Harmonic and Individual Lines plus Noise* («гармонические и индивидуальные составляющие плюс шум»), предназначены для кодирования неречевых сигналов со скоростями от 4-х кбит/с и выше.

Указанный набор средств может выбираться вручную: либо только *HVXC* либо только *HILN*. В этом случае выбранный режим используется для всех кодируемых аудиосигналов.

Интегрированный параметрический кодер при кодировании звукового сигнала использует средства *HVXC* и *HILN* поочередно или одновременно. Такой кодер автоматически использует то средство кодирования, которое наилучшим образом подходит к текущим характеристикам исходного сигнала. При этом для речевого сигнала используется режим *HVXC*, а для музыкального – режим *HILN*.

Выбор режима работы и, следовательно, используемых для кодирования текущей выборки средств, делается автоматически с помощью устройства классификации речь/музыка. Для сигналов, представляющих собой смесь речи и музыки, возможно одновременное использование средств *HVXC* и *HILN*.

Устройство классификации речь/музыка. Данное устройство принимает решение на основе анализа текущих характеристик звукового сигнала. При этом оцениваются энергия основного тона и энергия сегмента сигнала в целом.

В общем случае речь имеет большую интенсивность основного тона и более частое и большее изменение энергии сигнала в пределах сегмента, чем музыка.

Классификация сигнала речь/музыка может выполняться двумя способами:

- в первом из них анализируются первые 5 секунд кодируемого сигнала, и затем в соответствии с принятым в результате этого анализа решением для кодирования выбирается средство *HVXC* или *HILN*;

-во втором случае устройство классификации работает постоянно, его текущее решение используется для выбора *HVXC* или *HILN* при кодирования текущего сегмента. При этом нужно принимать во внимание, что задержка в принятии решения равна 5 с.

Оценка энергии сегмента кодируемого сигнала вычисляется по формуле:

$$\mathring{A}_{\tilde{n}\mathring{a}\tilde{a}\ddot{i}} = \sum_{n=0}^{159} s(n)^2 ,$$

где *s(n)* – отсчеты входного сигнала, *n* – номер отсчета. В этом случае используются сегменты с уровнями энергии, превышающими предварительно определенный минимальный уровень (> -78 дБ). Кратковременная средняя энергия сегмента определяется как среднее значение энергий четырех последних сегментов

$$Eav = \sum_{t=0}^{3} E_{сегм}\{t\} / 4 , \qquad t \text{ – номер сегмента.}$$

Далее вычисляется разность между энергией сегмента и средней кратковременной энергией сегмента

$$Ed[frm] = \left| E_{сегм} - Eav \right| / Eav .$$

Вычисленные значения *Ed[frm]* сохраняются в памяти примерно для 250 сегментов, что соответствует длительности звукового сигнала равной 5 с.

Оценка энергии основного тона. В кодере *HVXC* максимальная автокорреляция *LPC*-остатка (r0r) вычисляется в процессе определения основного тона. Значения r0r сохраняются примерно для 250 сегментов.

Принятие решение речь/музыка. Среднее значение $\mathring{A}d(av)$, отклонение энергии $\mathring{A}d(va)$ сигнала сегмента, а также соответствующие средние значения $r0r(av)$ и отклонение $r0r(va)$ величины *r0r* вычисляются соответственно как

$$Ed(av) = \sum_{frm=0}^{249} Ed[frm]/250,$$

$$Ed(va) = \sqrt{\sum_{frm=0}^{249} (Ed[frm] - Ed(av))^2 / 250},$$

$$r0r(av) = \sum_{frm=0}^{249} r0r[frm]/250,$$

$$r0r(va) = \sqrt{\sum_{frm=0}^{249} (r0r[frm] - r0r(av))^2 / 250}.$$

Речевые данные имеют большие отклонения, чем музыкальные в том же диапазоне средней величины *r0r*.

Совокупность полученных значений разделяется на три области:

(1) речь, если $r0r(va) \geq 1{,}153 r0r(av) + 0{,}113$;

(2) неизвестный сигнал, если

$$0{,}07 r0r(av) + 0{,}137 < r0r(va) < 0{,}153 r0r(av) + 0{,}113;$$

(3) музыка, если $0{,}07 r0r(av) + 0{,}137 \geq r0r(va)$.

Если среднее и отклонение лежат в области (1), то данные классифицируются как речь. Если они находятся в области (3), то классифицируются как музыка.

Если среднее и отклонение попадают в область (2), то дополнительно используются среднее и отклонение (дифференциальной) энергии сегмента *Ed*. Речевые данные имеют большие средние и отклонения *Pd*, чем музыкальные данные. В этой ситуации речевые и музыкальные данные разделяются в соответствии с нижеприведенными неравенствами:

речь: $Ed(va) \geq -0{,}5 Ed(av) + 0{,}8$,

музыка: $Ed(va) < -0{,}5 Ed(av) + 0{,}8$.

Режимы работы интегрированного параметрического кодера. Данный кодер может работать в следующих режимах:

Индекс режима работы кодера	Режим работы кодера
0	Только *HVXC*
1	Только *HILN*
2	Переключение *HVXC/HILN*
3	Смешанный *HVXC/HILN*

Режимы работы 0 и 1 представляют фиксированные режимы *HVXC* или *HILN*. Режим 2 позволяет осуществлять переключение между *HVXC* и

HILN в зависимости от текущего типа входного сигнала. В режиме 3 кодеры *HVXC* и *HILN* могут использоваться одновременно, а их выходные сигналы складываются (смешиваются) в декодере.

Интегрированный параметрический кодер обычно использует длину сегмента звукового сигнала равную 40 мс и частоту дискретизации 8 кГц. Он может работать на скорости равной или превышающей значение 2025 бит/с, но не более 4-х кбит/с.

Режим переключения HVXC/HILN. Устройство классификации *речь/музыка* основано на *HVXC*-кодере. Поэтому *HVXC*-кодер работает непрерывно для каждого сегмента звукового сигнала. Поток битов, сгенерированный *HVXC*-кодером, и входной звуковой сигнал сохраняются в двух буферах *FIFO* для компенсации 5-ти секундной задержки в принятии решения *речь/музыка*. Если сегмент сигнала классифицируется как речь, то бит «*PARAswitchMode*» устанавливается в *0*, и сигнал с выхода буфера *FIFO* передается *HVXC*-кодеру. В случае решения «музыка», когда бит *PARAswitchMode* устанавливается в *1*, сигнал с выхода буфера *FIFO* кодируется кодером *HILN*, и именно этот поток битов передается. Если для кодирования сегмента сигнала используется *HVXC*-кодер, то *HILN*-кодер отключается (*prevNumLine = 0*).

Смешанный режим HVXC/HILN. При работе параметрического кодека в смешанном режиме компоненты речь и музыка предварительно должны быть разделены. В этом случае процедура кодирования является наиболее простой в реализации.

2.5. Алгоритм кодирования *MPEG*-4 *SBR*

Иногда, например, в системе цифрового радиовещания DRM, при кодировании ЗС сложной структуры для большего снижения скорости цифрового потока дополнительно используется алгоритм, называемый *SBR* (*Spectral Band Replication*). Он является дополнением к стандарту *MPEG*-4. Существуют два различных протокола кодирования цифровых ЗС, предусматривающих совместное использование методов *SBR* и *MPEG*-4: *SBR* и *MPEG*-4 *AAC*; *SBR* и *MPEG*-4 *CELP*. Алгоритм кодирования *CELP* рассмотрен в следующем разделе.

Известно, что подавление высокочастотных составляющих в спектре ЗС (рис. 2.16.) приводит к искажению его тембра. Эта процедура часто имеет место при кодировании ЗС с малой скоростью цифрового потока, когда высокочастотные компоненты ЗС не кодируются из-за малого числа доступных бит. Тембр звука становится более глухим и тусклым, а звуковой сигнал, кроме того, – менее разборчивым и прозрачным, исчезают присущие ему тонкие детали, подчеркивающие индивидуальность звучаний музыкальных инструментов и голосов.

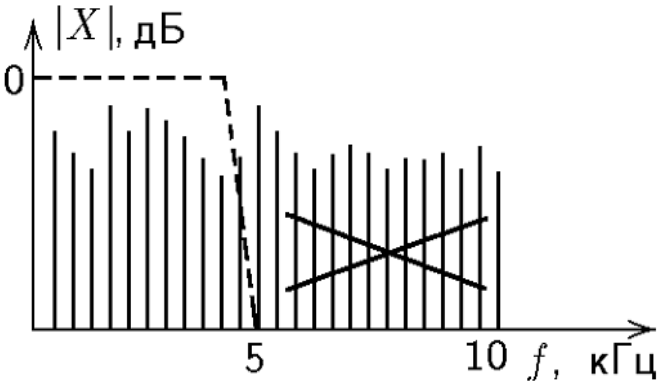


Рис. 2.16. Пример подавления высокочастотных составляющих спектра звукового сигнала

Метод *SBR* позволяет расширить полосу воспроизводимых частот звукового сигнала. Он основан на том, что подавленные на передающей стороне (в кодере) высокочастотные составляющие спектра ЗС могут быть воссозданы на приемной стороне (в декодере) при использовании дополнительной информации, выделенной на передающей стороне *SBR*-кодером. Возможны две версии алгоритма *SBR*: среднего (*Low Power*) и высокого (*High Quality*) качества.

Базовая структурная схема *SBR*-кодера представлена на рис. 2.17. Вся процедура обработки аудиосигнала с использованием данных *SBR*-кодека показана на рис. 2.18. Она не требует дополнительных пояснений.

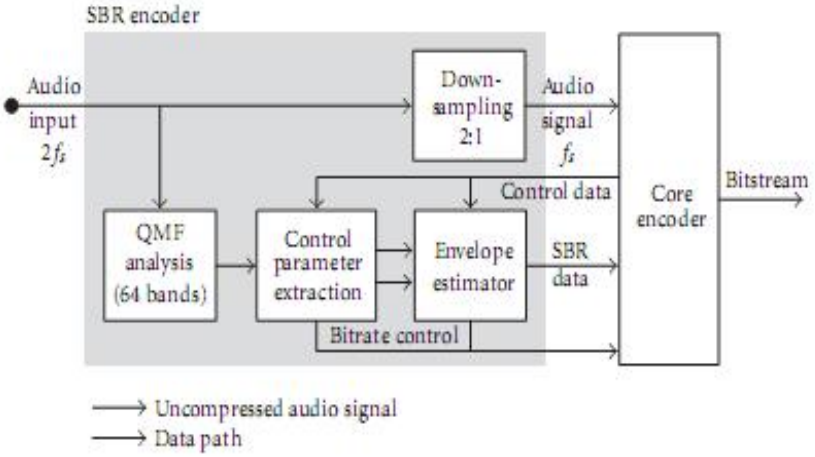


Рис. 2.17. Базовая структурная схема кодера SBR (дополнение к стандарту MPEG-4)

В *SBR*-кодере (*SBR encoder*, рис. 2.18) звуковой сигнал разделяется банком полифазных *QMF*-фильтров на 64 субполосные компоненты, каждая из которых имеет полосу частот 344 Гц при частоте дискретизации 44,1 кГц,

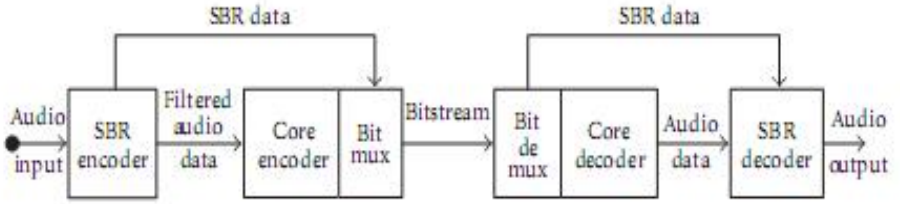


Рис. 2.18. Структурная схема кодека SBR

что соответствует временному разрешению 1,4 мс. Низкочастотная часть исходного звукового сигнала в полосе частот от 0 до 5,5 кГц кодируется обычно кодером *AAC* (на рис. 2.18, это *Core encoder*). Однако, в принципе, для ее кодирования может использоваться любой алгритм группы стандартов *MPEG*, например, *MPEG Layer* 2 или *MPEG Layer* 3, известный как mp3PRO. Высокочастотная часть ЗС обрабатывается *SBR* - кодером. Анализирующий полифазный банк *QMF* - фильтров кодера *SBR* разделяет исходный звуковой сигнал на комплексные субполосные составляющие, каждая из которых содержит вещественную и мнимую компоненты (рис. 2.19).

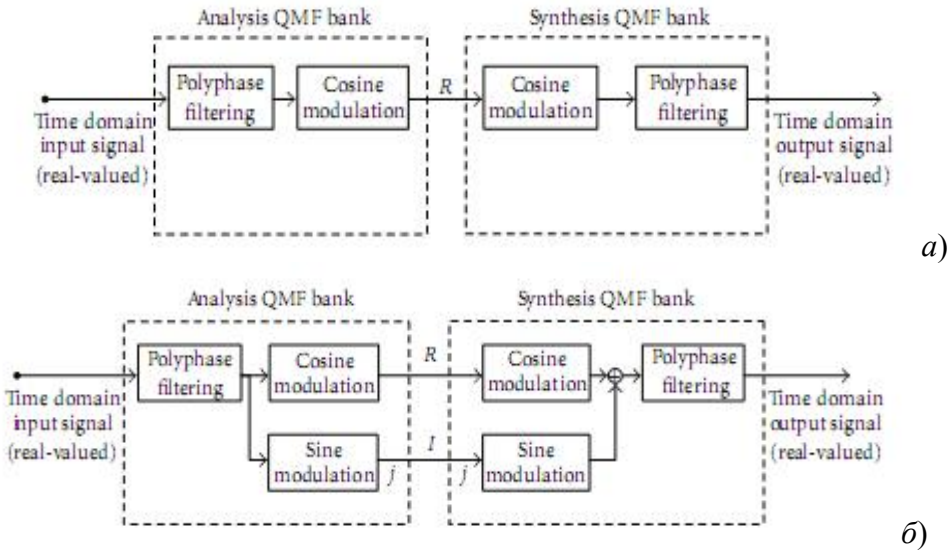


Рис. 2.19. Упрощенная структурная схема анализирующего и синтезирующего банков *QMF*-фильтров *SBR*-кодека: а – вещественная фильтрация; б – комплексная фильтрация

Работа с комплексныи сигналами позволяет упростить ряд вычислений. На следующем этапе, выполняемом в блоке выделения контрольных параметров кодера для их последующего включения в цифровой поток данных *SBR*, формируется частотно-временная *сетка*. Пример такой сетки представлен на рис. 2.20. Здесь по вертикальной оси отложена частота в кГц, а по гризонтальной оси – время в с. Структура данной сетки (длительность входящих в нее сегментов) может меняться от одной выборки сигнала к

другой. Белые пунктирные линии показывают на рис. 2.20 границы время-частотных областей, в пределах которых пррисходит выделение данных

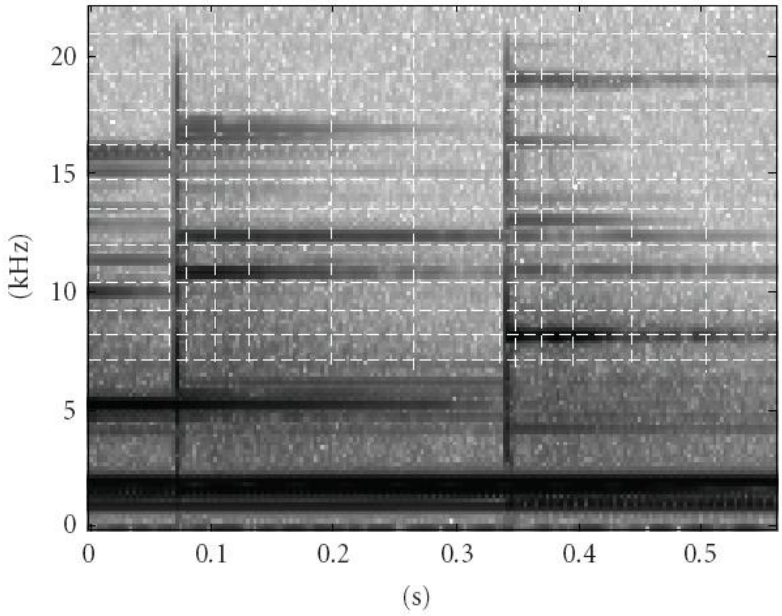


Рис. 2.20. Пример время-частотной сетки *SBR*-кодера (границы время-частотных областей для кодируемого сегмента ЗС показаны белыми пунктирными линиями)

SBR. Для каждого сегмента данной сетки вычисляется энергия сигнала, несущая информацию об спектралной огибающей исходного сигнала. Длины сегментов могут меняться от одной выборки сигнала к другой. При этом длительные сегменты (длинные выборки) с большим разрешением по частоте используются при вычислении огибающей ЗС, когда нет резких изменений сигнала по амплитуде, а более короткие сегменты (короткие выборки) с меньшим разрешением по частоте при резких изменениях сигнала по амплитуде.

Для выделения резких энергетических всплесков в сигнале используется так называемый детектор выбросов. Как только детектор выбросов сигнализирует о наличии такого всплеска длина сегмента, который используется для вычисления огибающей, становится минимальной. Его начальная граница расположена в начале выброса. Более длинные сегменты, следующие вслед за коротким сегментом, позволяют более правильно оценить спад переходного процесса (выброса сигнала), и наконец, самые длинные сегменты используются для обработки стационарной части сигнала (рис. 2.20).

Итак, в кодере используются при оценке огибающей два типа выборок: длинная (с высоким разрешением по частоте) и короткая (с высоким раре-шением по времени).

Группирование субполосных сигналов по частоте может быть сделано с использованием либо линейной, либо логарифмической шкалы с перемен-

ным числом субполос на октаву. Можно сказать, что процесс группирования субполосных компонент и длины сегментов адаптированы к сигналу. При кодировании огибающей вычисляется энергия в каждом таком сегменте сигнала.

Кроме параметров огибающей звукового сигнала *SBR*-кодер (рис. 2.17) вычисляет также уровень дополнительного шума, который следует добавить в сгенерированную в декодере высокочастотную компоненту сигнала. Для этой цели используется технология «*анализ через синтез*». Суть этого процесса показана на рис. 2.21. По вертикальной оси отложен уровень ЗС в

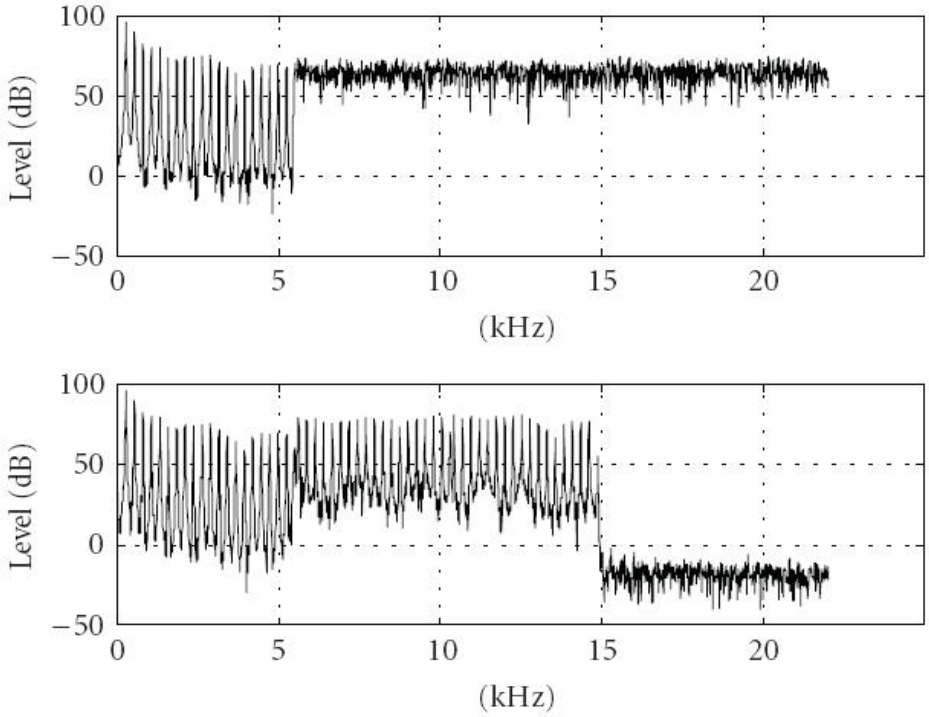


Рис. 2.21. Спектр декодированного сигнала: сгенерированная декодером низкочастотная часть входного сигнала (а); спектр сигнала, полученного в декодере после генерирования высокочастотной части (на базе его низкочастотной части), но без шумовой коррекции (б)

дБ, по горизонтальной оси – частота в кГц. Здесь на верхнем графике изображен спектр исходного звукового сигнала, который резко обрывается на частоте 5,5 кГц. Остальной спектр данного сигнала – это шум. Напомним, что в низкочастотной области звуковых сигналов расположены основой тон, гармонические компоненты кратные по частоте сигналу основного тона и тональные компоненты высокого уровня, очень важные для слухового восприятия. На более высоких частотах спектр звукового сигнала обычно имеет шумовую структуру. Число спектральных компонент высокого уровня здесь крайне мало. На нижнем графике (рис. 2.21) показан этот же сигнал, но для него в декодере сгенерирована высокочастотная компонента без

какой-либо коррекции огибающей. В этом простом случае высокочастотная часть сгенерированого сигнала в полосе частот 5,5…15 кГц будет сдержать большое число спектральных компонент высокого уровня, что явно не соответствует спектру исходного звукового сигнала в этой части. Добавление белого шума позволяет скорректировать это несоответствие. Процедура вычисления уровня добавочного шума (второй группы параметров, передаваемых кодером *SBR* - декодеру) требует дополнительного изучения.

Кодеру *SBR* требуется также оценить: не будут ли при этом потеряны тональные компоненты высокого уровня, имеющиеся в высокочастотной части исходного звукового сигнала. В качестве примера на рис. 2.22 показан такой пример, когда три спектральные компонены высокого уровня,

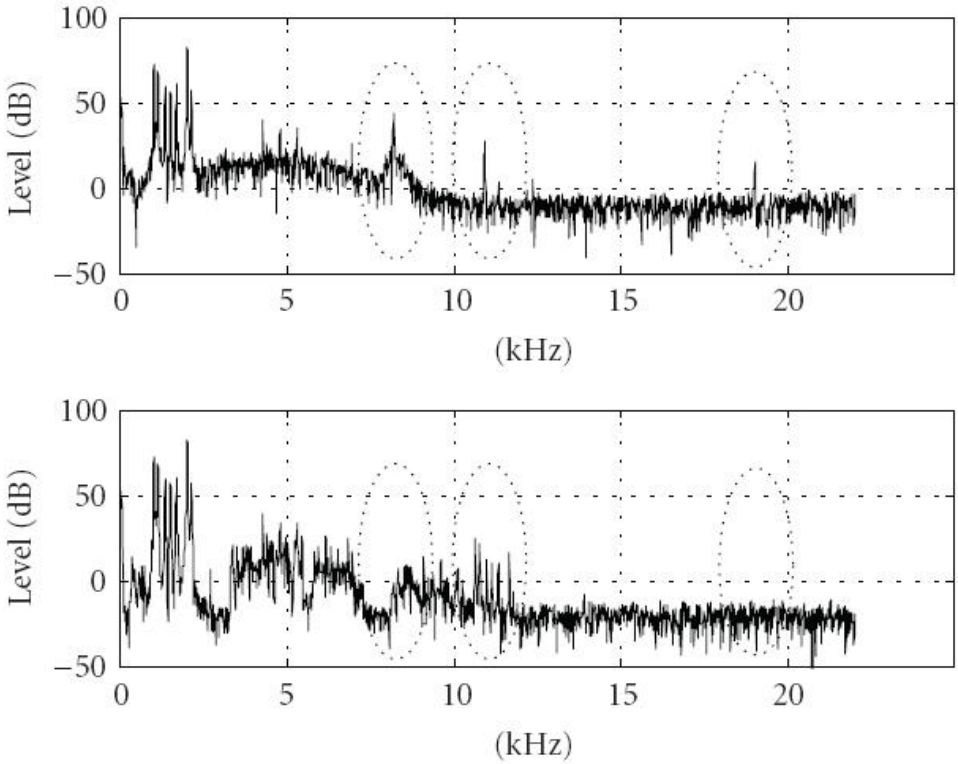


Рис. 2.22. Примеры спектров: исходный звуковой сигнал с тремя тональными компонентами высокого уровня в высокочастотной части (*а*, выделены овалами) и их отсутствие при реконструкции в декодере (*б*), когда в кодере не использовалась процедура выделения, квантования и кодирования их параметров, а в декодере не выполнена их реконструкция по переданным от кодера параметрам

имеющиеся в высокочастотной части спектра исходного сигнала, не восстановились после его реконструкции в декодере, когда выполнялась только процедура коррекции огибающей. Спектральные компоненты выского уровня выделены здесь овалами. Очевидно, что они долны быть также сгенерированы в декодере. Чтобы это стало возможным, информация о часто-

те и уровне этих спектральных компонент высокого уровня должна быть передана декодеру, т.е. включена в цифровой поток данных *SBR* - кодера.

Информация об огибающей кодируемого звукового сигнала, уровне добавочного шума и спектральных компонентах высокого уровня в высокочастоной части спектра ЗС квантуется и кодируется кодером *SBR*. При этом кодируются их дифференциальные значения с использованием таблиц кодов Хаффмана. Иначе говоря, здесь применяется энтропийное кодирование дифференциальных значений данных параметров кодовыми словами переменной длины.

Для обеспечения кодирования выбросов ЗС независимо от их расположения внутри аудиофрейма сами фреймы имеют разную длину, дргими словами их длина может меняться при кодировании.

Базовая структура декодера *SBR* представлена на рис. 2.23. В основном

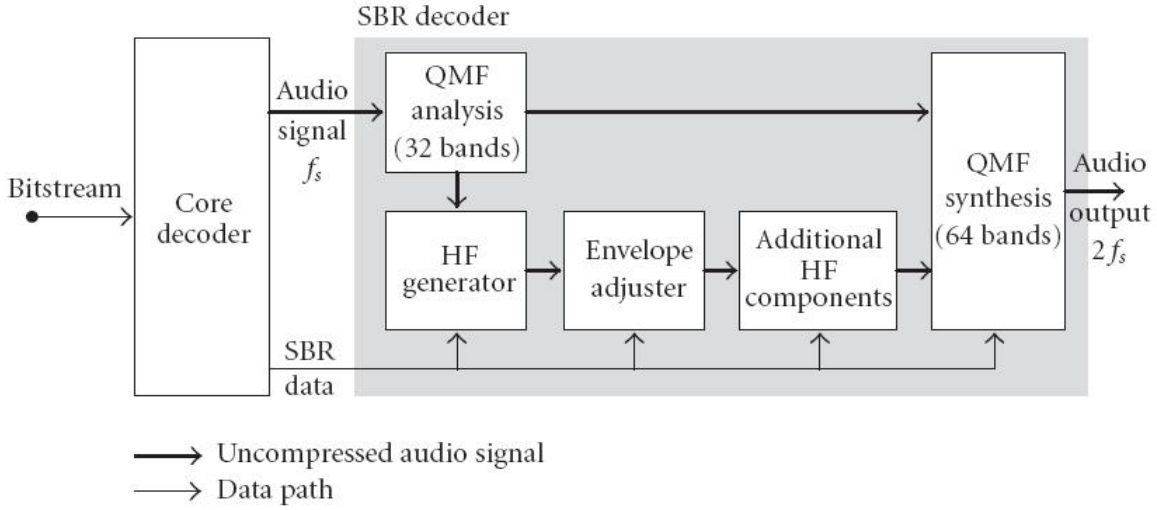


Рис. 2.23. Базовая структурная схема SBR-декодера

декодере входной цифровой поток разделяется на две части: цифровой поток данных *SBR* и цифровой поток кодера *AAC*. После декодирования последнего получается низкочастотная часть спектра исходного звукового сигнала. Далее этот реконструированный сигнал поступает на банк анализирующих полифазных *QMF* -фильтров. Он разделяет низкочастотную часть сигнала на 32 субполосные составляющие, каждая из которых имеет соответственно вещественную и мнимую части с соотвествующим понижением частоты дискретизации в каждом таком субполосном канале. В каждой из 32 субполос образуются группы по 30 отсчетов ЗС. В результате на выходе анализирующего банка фильтров формируется аудиофрейм, содержащий в общей сложности 960 отсчетов. Эти фреймы поступают на устройство задержки, которое необходимо для согласования по времени сигналов низкочастотных и высокочастотных субполос, и на устройство вос-

создания высокочастотных спектральных составляющих (ВЧ-генератор). На это же устройство поступает необходимая информация с блока деформатирования цифрового потока декодера *SBR*. Декодер Хаффмана преобразует принятые кодовые слова в квантованные отсчеты огибающей ЗС, дополнительного шума и тональные компоненты высокого уровня для их последующего введения в спектр сгенерированной высокочастотной части ЗС.

Комплексные субполосные составляющие ВЧ-генратором переносятся в область высоких частот, создавая, таким образом, высокочастотную часть спектра исходного сигнала. Генерирование высокочастотных субполосных сигналов выполняется выбором определенных полос низкочастотного сигнала по определенному правилу. Алгоритм переноса (копирования) низкочастотной части сигнала в высокочастотную область 5,5…15 кГц учитывает следующие факторы:

-переносимые полосы частот низкочастотной части должны покрывать диапазон частот до 16 кГц без использования самой низкой по частоте субполосной компоненты исходного низкочастотного сигнала, включая постоянную составляющую;

-на более низких частотах чувствительнось слуха выше, поэтому группы субполосных составляющих, имеющие более широкий спектр, следует переносить в более низкочастотную часть генерируемой высокочастной части спектра звукового сигнала, чтобы переместить возможный разрыв между первым и вторым участком в область более высоких частот;

-временная структура сигнала высокочастотной части должна быть по возможности ближе к структуре низкочастотной части, что позволяет добиться лучше сходства с оригиналом на следующих этапах ее обработки;

-при простом копировании низкочастотной части сигнала в высокочастотную часть последняя будет содержать значительное число высоких по уровню тональных компонент, это требует дополнительной ее обработки с тем, чтобы уменьшить их влияние при слуховом восприятии.

Для уменьшения «*тональности*» высокочастотной части (для ее «исправления») реконструируемого сигнала к сгенерированным высокочастотным субполосным сигналам применяется процедура линейного предсказания второго порядка, реализованная на основе адаптивных FIR-фильтров низкого порядка. Коэффициенты предсказания фильтров определяются на основе анализа спектра низкочастотного сигнала и цифровых данных SBR-потока.

Пример генерирования (копирования) низкочастотной части ЗС в высокочастоные области (А и В) показан на рис. 2.24. При этом область частот А несколько шире по полосе области В, ибо верхняя частота спектра реконструированного сигнала ограничена частотй 15 кГц. Самая нижняя субполоса низкочастотной части ЗС не копируется (не переносится) в область высоких частот. Излишняя «*тональность*» реконструированного

сигнала областях 1, 2 и 3 устраняется с помощью процедуры линейного предсказания второго порядка, выполняемой в каждой субполосе.

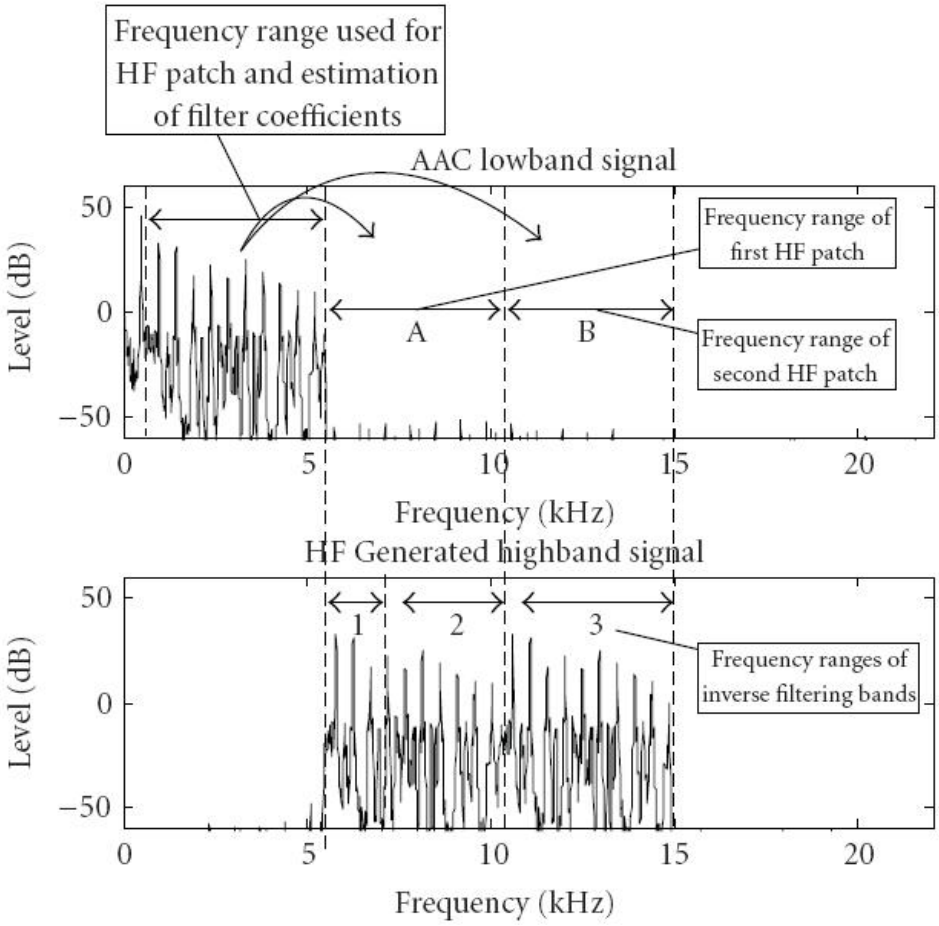


Рис. 2.24. Пример генерирования (копирования) низкочастотной части звукового сигнала в высокочастотные области (А и В), полоса частот от 5,5 до 15 кГц. Сигнал представляет собой отрывок классической музыки, моно, скорость цифрового потока 24 кбит/с

Сгенерированные и исправленные высокочастотные сигналы затем подаются в блок коррекции огибающей. Это самая большая часть потока данных *SBR*. После коррекции огибающей в полученный сигнал вводятся дополнительные спектральные компоненты высокого уровня, сведения о которых также содержатся в цифровом потоке *SBR*-кодера.

При этом важно поддержание соотношений между гармоническими и шумоподобными компонентами в воссозданной высокочастотной части спектра ЗС. После реконструкции высокочастотной части исходного сигнала обе его компонены объединяются в синтезирующем полифазном *QMF*-фильтре. Банк синтезирующих фильтров декодера также комплексный, но мнимая компонента сигнала на его выходе отбрасывается.

Информация о частотных диапазонах и временных интервалах, действительных для каждого фрейма (частотно-временные параметры), передается на декодер.

Границы временных интервалов выбираются в соответствии со свойствами ЗС. Более длинные интервалы используются для квазистационарных ЗС, а более короткие – для быстро изменяющихся звуковых сигналов. Временные и частотные параметры, определяющие шумоподобные спектральные составляющие ЗС, передаются аналогичным образом.

На приемной стороне информация с выходов декодера Хаффмана и устройства управления частотно-временными параметрами поступает на вход блока расчета коэффициентов усиления. Эти коэффициенты необходимы для формирования огибающей высокочастотной части спектра ЗС в блоке регулировки усиления.

В версии *LP*, когда используются вещественные банки *QMF*-фильтров (рис. 2.19,*а*), возможно появление в субполосных сигналах дополнительных компонент, обусловленных эйлайзингом при коррекции огибающей. Пример коррекции огибающей при использовании банка вещественных *QMF*-фильтров показан на рис. 2.25. На верхнем рисунке изображены тональные компоненты внутри субполос банка *QMF*-фильтров. Здесь же пун-

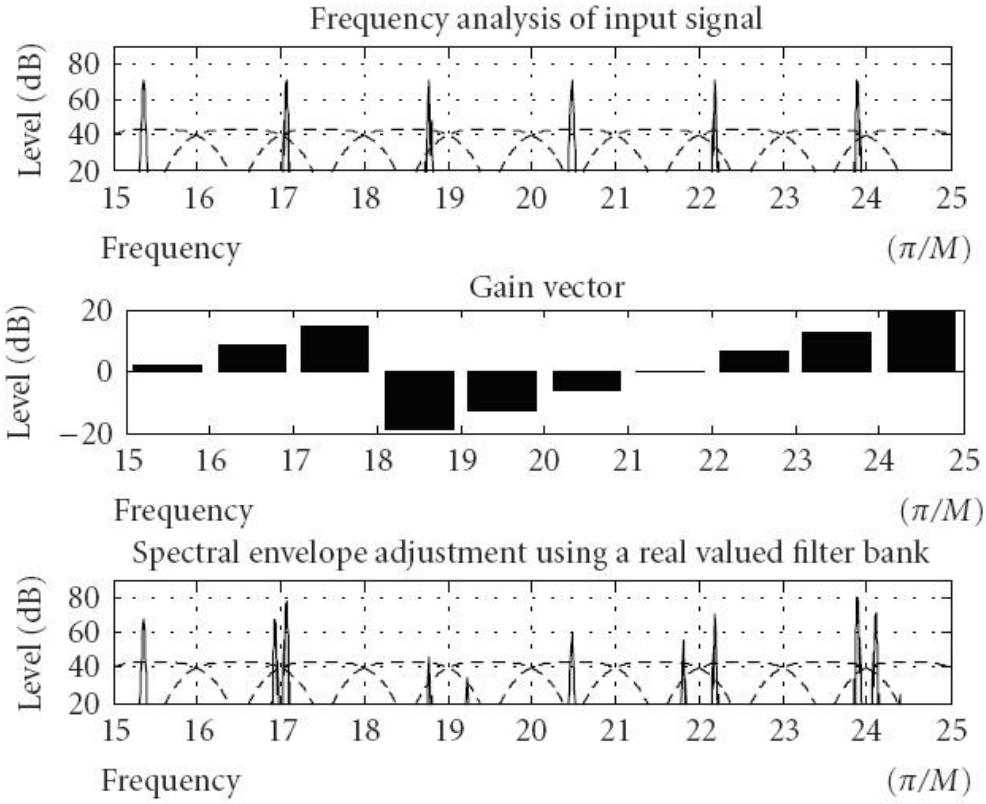


Рис. 2.25. Пример коррекции огибающей в наборе вещественных *QMF*-фильтров

ктиром показаны частотные характеристики самих фильтров, причем тональные компоненты расположены в местах перекрытия субполос. На

среднем рисунке изображены условно в форме черных прямругольников коэффициенты усиления (в дБ) в субполосах, вычисленные блоком коррекции огибающей. На нижнем графике показаны скорректированные по амплитуде исходные тональные компоненты с учетом рассчитанных коэффициентов усиления в субполосах, а также здесь видны дополнительные тональные компоненты, появившиеся в выходном сигнале вследствие явления эйлайзинга и отсутствующие в исходном сигнале. На рис. 2.26 показано декодирование этого сигнала, позволяющее исключить появление последних.

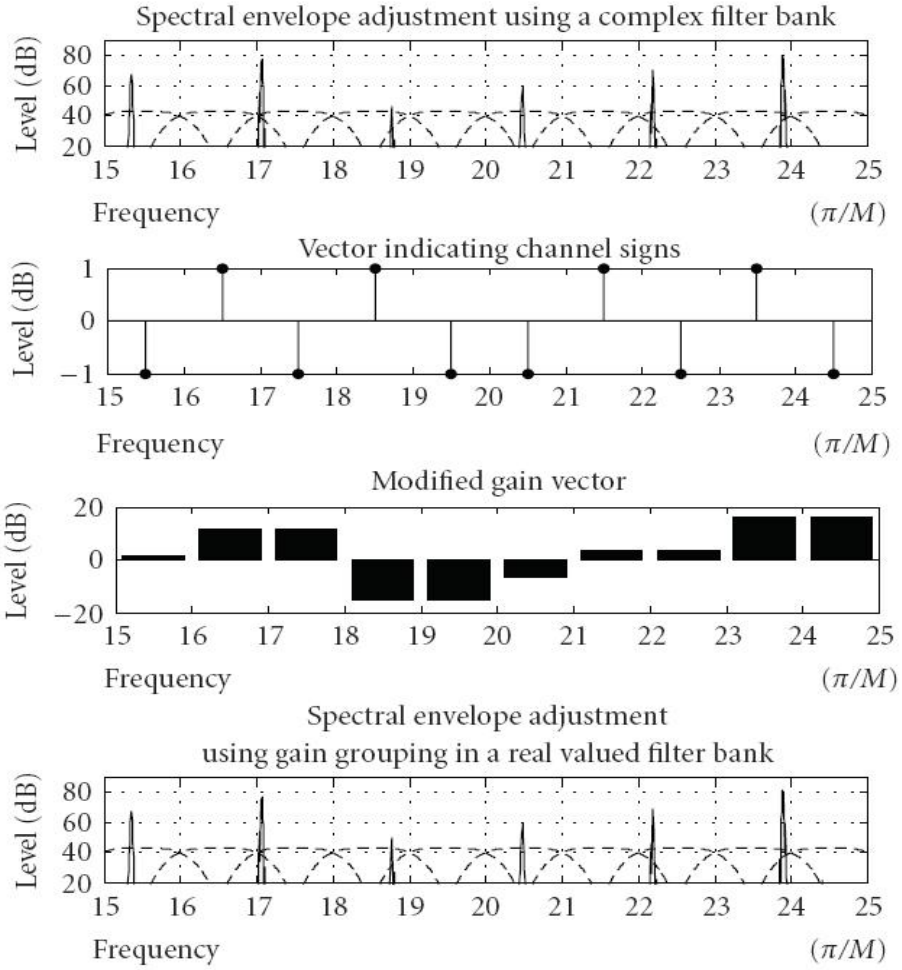


Рис. 2.26 Пример коррекции огибающей в наборе вещественных *QMF*-фильтров с устранением явления эйлайзинга

Этот рисунок отличается от предшествующего графика «знаками коэффициентов усиления в субполосных каналах». Эти знаки получены из выражения

$$\mathrm{Sign}(x) = \begin{cases} (-1)^{k}, \text{ если } a_1 < 0 \\ (-1)^{\kappa+1}, \text{ если } a_1 > 0, \end{cases}$$

где коэффициент a_1 получен из формулы для передаточной функции фильтра линейного предсказания первого порядка для соответствующей субполосы,

$$A(z) = 1 - a_1 z^{-1},$$

к – номер субполосы, рассчитывамый от 0. Для соседних субполос, где низкая по частоте субполоса имеет знак плюс, а верхняя – минус, коэффициенты фильтра предсказания должны рассчитываться зависимо. Для других ситуаций они могут быть рассчитаны независимо. После этой дополнительной коррекции коэффициенты усиления в соседних субполосах с общими тональными компонентами оказываются одинаковыми (третий график сверху на рис. 2.26). В этом случает дополнительные тональные компоненты, обусловленные эйлайзингом, не появляются (нижний график на рис. 2.26).

Спектры сигнала на разных этапах его обработки в декодере показаны на рис. 2.27. На левом верхнем рисунке изображен спектр исходного звукового сигнала; на правом верхнем – спектр низкочастотной части исходного

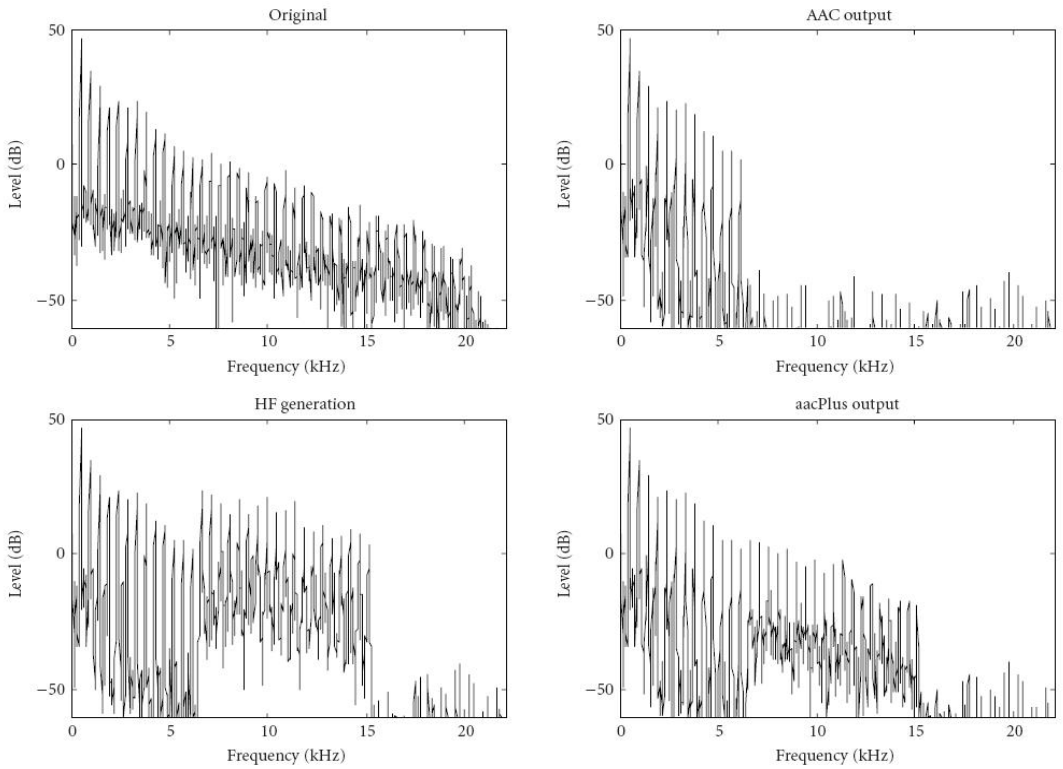


Рис. 2.27. Спектр звукового сигнала на разных этапах его обработки: левый верхний – энергетический спектр исходного звукового сигнала; правый верхний – спектр низкочастотной части сигнала на выходе декодера *AAC*; левый нижний – спектр низкочастотной и высокочастотной частей сигнала перед коррекцией огибающей; правый нижний – спектр выходного сигнала после коррекции огибающей его высокочастотной части

сигнала после его декодирования в основном декодере (рис. 2.23); на левом нижнем – спектр низкочастной части декодированного сигнала и реконструированная его высокочастотная часть без коррекции огибающей; правый нижний – спектр выходного сигнала декодера.

Синтезирующая фильтрация задержанных отсчетов низкочастотных субполос и высокочастотных субполосных отсчетов, прошедших процедуру регулировки усиления, выполняется при помощи 64 канального банка фильтров. Отсчеты низкочастотных субполос поступают на низшие 32 канала синтезирующего банка фильтров, а высокочастотных – на остальные 32 канала, соответствующие высоким частотам. Аудиофрейм, сформированный на выходе синтезирующего фильтра, содержит 1920 отсчетов ЗС и состоит из двух частей, относящихся к *MPEG*-4 *AAC* и SBR, соответственно (рис. 2.28).

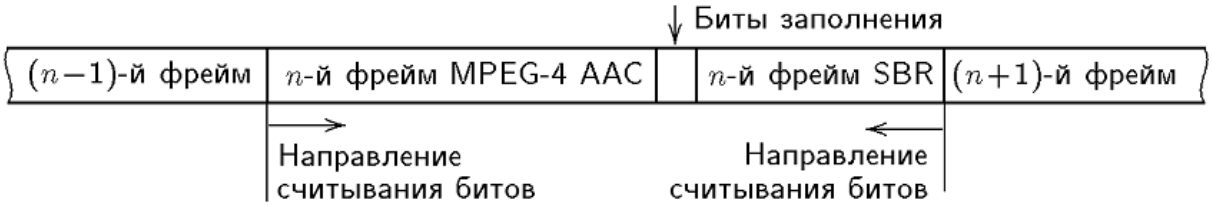


Рис. 2.28.Структура данных аудиофрейма с данными *SBR*-кодера

Биты SBR-данных расположены в конце фрейма. Направления считывания битов в частях, относящихся к *MPEG*-4 *AAC* и *SBR*, противоположны, что облегчает поиск стартовых позиций обеих частей фрейма.

Эффективность метода *SBR* можно оценить на представленном ниже примере кодирования монофонического ЗС. Для этого случая получены следующие данные.

Скорость передачи ЗС	22 кбит/с .
Длительность аудиофрейма	40мс.
Частота дискретизации *MPEG* - 4 *AAC*	24 кГц.
Частота дискретизации *SBR*	48 кГц
Частотный диапазон ЗС при применении *MPEG* - 4 *AAC*	0-6 кГц.
Частотный диапазон ЗС за счет применения *SBR*	6-15,2 кГц.
Средняя скорость цифрового потока *SBR* на канал	2 кбит/с

При кодировании ЗС методом *MPEG* - 4 *AAC* в данном случае можно обеспечить диапазон воспроизводимых частот от 0 до 6кГц. Применение дополнительно метода *SBR* позволяет расширить диапазон воспроизводи-

мых частот с 6 кГц до 15,2 кГц. При этом общая скорость передачи цифрового потока составляет примерно 22 кбит/с.

В заключение приведем результаты тестирования (рис. 2.29) качества

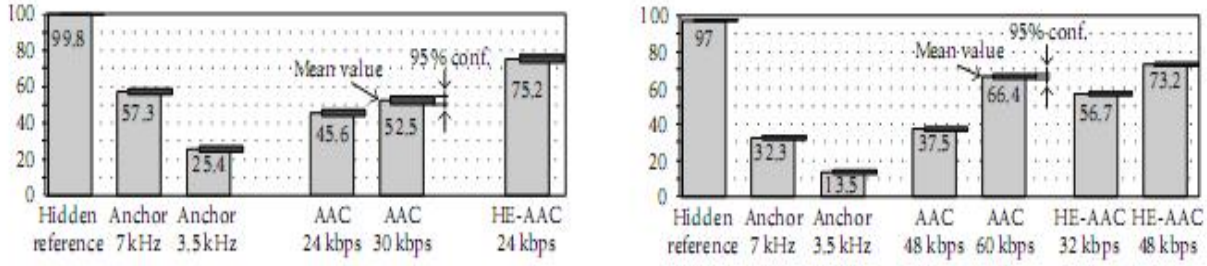


Рис. 2.29. Результаты тестирования кодеков стандарта *MPEG* - 4 *AAC* и *MPEG*-4 *HE+SBR* в шкале *MUSHRA* для разных значений скорости цифрового потока: а – монофонические испытательные сигналы; б – то же самое, но при стереофоническом тестовом сигнале. Здесь представлены результаты, усредненные для звуковых сигналов разных жанров

кодеков *MPEG* -4 *AAC* и *MPEG* - 4*HE-AAC* (*AAC+SBR*), выполненные методом субъективно-статистических экспертиз в шкале оценки *MUSHRA* – *Mtthod for subjective assessment of intermediadate quality level of coding systems*, *ITU-R Recommend*. *BS*.1534. Видно (рис. 2.29), что технология кодирования *AAC+SBR* обеспечивает более высокое качество при меньшей скорости цифрового потока.

2.6. Алгоритм *CELP* стандарта *MPEG* - 4

Общая часть

Метод кодирования *MPEG* - 4 *CELP* стандарта *ISO*/*IEC* 14496-3 предназначен для обработки речевых сигналов (РС). Заметим, что устройства кодирования речи можно разделить на две группы: кодеры формы сигнала и вокодеры. На практике применяются, в основном, три основных класса кодеров: кодеры формы, вокодеры и гибридные кодеры. Заметим, что данный раздел написан по результатам работ проф. Л.Н.Кацнельсона.

Кодеры формы характеризуются способностью сохранять основную форму речевого сигнала. К ним относятся кодеры с импульсно-кодовой модуляцией (ИКМ), кодеры с дифференциальной ИКМ (ДИКМ), адаптивной дифференциальной ИКМ (АДИКМ) и др. Системы передачи с подобным типом кодеров обеспечивают хорошее качество воспроизведения речевых сигналов (стандартная полоса частот которых составляет 300–3400 Гц) и более широкополосных звуковых сигналов. Однако, они малоэффективны, если говорить о снижении скорости передачи цифровых ау-

диоданных. Так, стандартный телефонный речевой сигнал в системе с ИКМ и мгновенным компандированием передается со скоростью 64 кбит/с. Применение АДИКМ позволяет снизить скорость передачи такого сигнала при сохранении приемлемого качества воспроизведения речи до 32 кбит/с, то есть всего в 2 раза.

Вокодеры (от английских слов «*voice*» - голос и «*coder*» - кодирующее устройство) обеспечивают значительно большее снижение скоростей передачи РС. Сжатие информации на передающей стороне производится в анализаторе, выделяющем из речевого сигнала медленно меняющиеся составляющие, которые передаются по каналу связи в виде кодовых комбинаций.

На приемной стороне имеются местные источники сигналов. Управление ими осуществляется на основе информации, содержащейся в указанных кодовых комбинациях. В результате синтезируется речевой сигнал.

Работа вокодеров основана на моделировании звуков речи с учетом ее характерных особенностей. Вокодер преобразует входной сигнал в некий другой, похожий на исходный. При этом измеряемые характеристики используются для подстройки параметров вокодера в соответствии с принятой моделью речевого сигнала. Именно эти параметры и передаются на декодер приемника, который по ним восстанавливает (синтезирует) речевой сигнал. При этом оценка качества воспроизведения речи (разборчивость, естественность, узнаваемость и др.) обычно производится с применением субъективно-статистических экспертиз.

Наибольшее распространение получили параметрические вокодеры, в которых из речевого сигнала выделяют два типа параметров:

параметры, характеризующие огибающую спектра речевого сигнала (фильтровую функцию);

параметры, характеризующие источник речевых колебаний (генераторную функцию): частоту основного тона, ее изменения во времени, моменты появления и исчезновения основного тона, шумового сигнала и др.

По этим параметрам на приемной стороне синтезируют речь.

Вокодеры с линейным предсказанием (*LPC – Linear Predictive Coding*)

В вокодерах с линейным предсказанием при анализе речевого сигнала в передающем устройстве (кодере) определяются коэффициенты предсказания, а в приемном устройстве (декодере) на основе этих коэффициентов с помощью рекурсивного цифрового фильтра синтезируется эквивалент голосового тракта.

Принцип метода линейного предсказания состоит в том, что прогнозируемая величина речевого сигнала $\overline{\lambda}(h)$ в момент выборки h определяется как линейно взвешенная сумма значений сигнала в моменты предшествующих выборок:

$$\overline{\lambda}(h)=\sum_{m=1}^{p}\overline{\lambda}(h-m)a_m\,,$$

где: $\overline{\lambda}(h\text{-}m)$ – значения речевого сигнала в моменты предшествующих выборок; m = 1,2…p; p – порядок предсказания; a_m – коэффициенты предсказания. Интервалы времени между моментами выборок определяются частотой дискретизации $t_h - t_{h\text{-}1} = 1/f_Д$. В момент h, когда известны не только предсказанное значения $\lambda(h)$, но и истинное значение речевого сигнала, можно определить ошибку предсказания $\varepsilon(h)=\lambda(h)-\overline{\lambda}(h)$ и затем подобрать коэффициенты предсказания таким образом, чтобы эта ошибка была минимальной.

Коэффициенты предсказания, значения которых передаются по каналу связи на приемную сторону, используются в качестве переменных параметров в рекурсивном цифровом фильтре, на вход которого подаются сигналы возбуждения. При воспроизведении вокализованных звуков (гласных) - это последовательность импульсов с частотой основного тона, а при воспроизведении невокализованных звуков (согласных) это случайная последовательность импульсов, формируемых генератором шума.

При кодировании с линейным предсказанием моделируются различные параметры человеческой речи, которые передаются вместо отсчетов речевого сигнала или их разностей. Это позволяет существенно снизить скорость передачи речевого сигнала по сравнению с методами ИКМ, ДИКМ, АДИКМ.

Широко применяемый в настоящее время метод кодирования с линейным предсказанием предусматривает формирование групп отсчетов, для каждого из которых вычисляется и передается частота основного тона, его амплитуда и информация о типе возбуждающего воздействия (гармоническое, негармоническое). Структура синтезатора речи с линейным предсказанием показана на рис. 2.30.

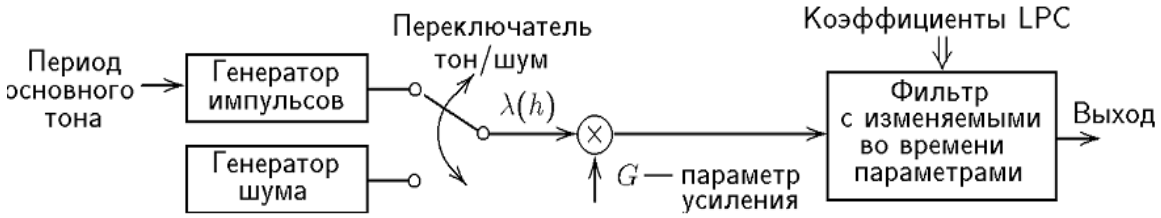


Рис. 2.30. Структура синтезатора речи с линейным предсказанием

Здесь сигналы возбуждения имеют вид последовательности импульсов на частоте основного тона (для вокализованных звуков) или случайного шума (для невокализованных звуков). Различные комбинации спектральных составляющих речи, образующейся, в частности, за счет работы голосовых связок, языка и губ человека, могут быть промоделированы цифро-

вым фильтром с изменяющимися параметрами. При линейном предсказании обычно производится спектральный анализ речи и выполняется построение систем анализа-синтеза. Во всех случаях параметры синтезатора обновляются при смене анализируемых кадров речевого сигнала. Чтобы избежать эффектов, связанных со скачками значений параметров, необходимо плавно их изменять с помощью интерполяции при переходе от одного фрагмента (сегмента) речи к другому.

При кодировании речевых сигналов по методу *LPC* обычно применяют метод анализа через синтез (*Analysis – by – Synthesis (AbS)*). При этом синтезатор (основной элемент декодера речевого сигнала) используется как составная часть кодера (рис. 2.31,*а*). На основе формируемых данных производится синтез речевого сигнала. Синтезированный речевой сигнал

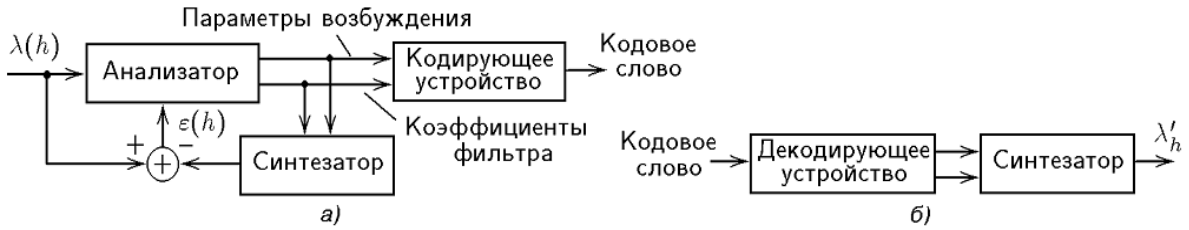


Рис. 2.31. Иллюстрация метода анализа через синтез: *а* — кодер; *б* — декодер

сравнивается в процессе передачи с реальным сигналом, поступающим на вход устройства. Сигнал ошибки *ε(h)*, получаемый в результате вычитания истинного и синтезированного сигналов, используется для повышения достоверности оценки кодируемых и передаваемых параметров. Структурная схема декодера представлена на рис. 2.31,*б*, где *(λ'(h)* – это значение речевого сигнала для момента времени *h*, полученное после декодирования).

По существу системы кодирования, использующие метод *LPC*, отличаются лишь способами генерирования возбуждающего воздействия и выбора параметров моделирующего фильтра.

Векторное квантование и кодовые книги

Когда набор значений амплитуд речевого сигнала, дискретизированного по времени, квантуется совместно как единый вектор, такой процесс называется векторным квантованием (*VQ – vector quantization*), известный также как блочное квантование [2.30].

Будем считать, что

$$\mathbf{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T$$

представляет собой N-мерный вектор с действительными значениями (символ T означает сдвиг по времени), а $\lambda_к$ – случайным образом меняющийся компонент с непрерывной амплитудой, где $1 \le к \le N$. При векторном квантовании вектору **Λ** ставится в соответствие другой N-размерный вектор **y**, с действительными значениями и дискретными амплитудами. Таким образом **Λ** квантуется как **y**. Другими словами, **y** используется для представления **Λ**.

Обычно **y** выбирается из конечного набора значений **Y**:

$$Y=y_i,\ 1 \le i \le L,\ \ y_i = [y_{i1}, y_{i2}, \ldots, y_{iN}]^T.$$

Набор значений **Y** называется кодовой книгой или шаблоном, L – размер кодовой книги, $\mathbf{y}_i$ - набор векторов кодовой книги [2.30].

Структурная схема простого векторного квантователя представлена на рис. 2.32. В линию связи передают только индексы i векторов $\mathbf{y}_i$, входящих в кодовую книгу. На приемной стороне имеется такая же кодовая книга. По принятому индексу i восстанавливают вектор $\mathbf{y}_i$.

Рис. 2.32. Структурная схема векторного квантования

Некоторые кодовые книги рассчитываются заранее и не изменяются. Они называются фиксированными кодовыми книгами. Другие кодовые книги могут обновляться в процессе работы. Одним из способов сделать кодовую книгу следящей за характеристиками входного вектора с течением времени является ее адаптация. Такие кодовые книги называются адаптивными. При обработке речевых сигналов применяются также случайные кодовые книги. Примером такой книги может быть гауссовская кодовая книга, которая содержит случайным образом выбранные векторы, сами содержащие случайные числа.

Векторное квантование может осуществляться не только с использованием значений амплитуд дискретизированного по времени сигнала. Многомерный вектор можно сформировать на основе гармонических составляющих спектра передаваемого речевого сигнала и создать соответствующую кодовую книгу, которая будет содержать конечное число значений такого вектора.

Метод кодирования *CELP*

Кодеры речевых сигналов, использующие алгоритм *CELP*, относятся к классу гибридных и занимают промежуточное положение между кодерами формы, в которых сохраняется форма колебания речевого сигнала в процессе его дискретизации и квантования, и параметрическими вокодерами, основанными на процедурах оценки и кодирования небольшого числа параметров речи. При этом в них сохраняются преимущества обоих типов кодеров.

Метод кодирования CELP основан на линейной авторегрессионной модели процесса формирования и восприятия речи и входит в группу методов *анализа через синтез*.

Линейная авторегрессионная модель процесса формирования речевых сигналов с локально постоянными на интервалах 10…30 мс параметрами, получившая в настоящее время широкое распространение, имеет вид [2.30]:

$$\lambda(h) = \sum_{m=1}^{M} a_m \lambda(h-m) + x(h),$$

где: M – порядок модели; $\lambda(h)$ – последовательность отсчетов речевого сигнала; a_m – коэффициенты линейного предсказания, характеризующие свойства голосового тракта; $x(h)$ – сигнал возбуждения голосового тракта (порождающая последовательность).

По существу, в алгоритме *CELP* производится векторное квантование порождающей последовательности $x(h)$. При этом отрезок (сегмент) сигнала возбуждения выбирается из предварительно сформированной совокупности кодовых комбинаций (векторов) кодовой книги, содержащей достаточно большое количество реализаций. В канал связи передаются индекс элемента кодовой книги с соответствующим коэффициентом усиления, параметры синтезатора основного тона, а также коэффициенты линейного предсказания, характеризующие состояние голосового тракта.

Авторегрессионная модель речевого сигнала описывает его с достаточно высокой точностью и позволяет применять для кодирования хорошо развитый математический аппарат линейного предсказания. Ее применение обеспечивает более высокое качество декодированной речи, устойчивость к входному акустическому шуму и ошибкам в канале связи по сравнению с иными принципами кодирования [2.30].

При использовании метода анализа через синтез задача анализа сводится к процедуре оценки передаваемых в канал связи параметров речи, проводимой в соответствии с некоторым критерием рассогласования между исходным и декодированным (синтезированным) сигналом. Метод *CELP*

эффективно применяется при кодировании речи со скоростями передачи от 4 кбит/с и выше.

Алгоритм *CELP* в системе *DRM*

В системе *DRM* применяется вариант 2 метода кодирования речи *MPEG*-4 *CELP* (стандарты *ISO/IEC* 14496-3 и *ISO/IEC* 14496-3/*Amd1*), обеспечивающий повышенную устойчивость к ошибкам (*Object Type ID* = 24, который является частью профиля *High Quality Audio Profile*).

Метод кодирования *CELP*, используемый в системе *DRM*, обеспечивает передачу речевых сигналов при скоростях цифровых потоков на выходах кодеров от 4 до 24 кбит/с. Для него в системе *DRM* предусмотрены два значения частоты дискретизации: $f_Д$ = 8 кГц и $f_Д$ = 16 кГц, что обеспечивает соответственно полосы звуковых частот равные 100…3800 Гц и 50…7000 Гц.

Базовая структурная схема декодера *MPEG*-4 *CELP* представлена на рис. 2.33 [2.31].

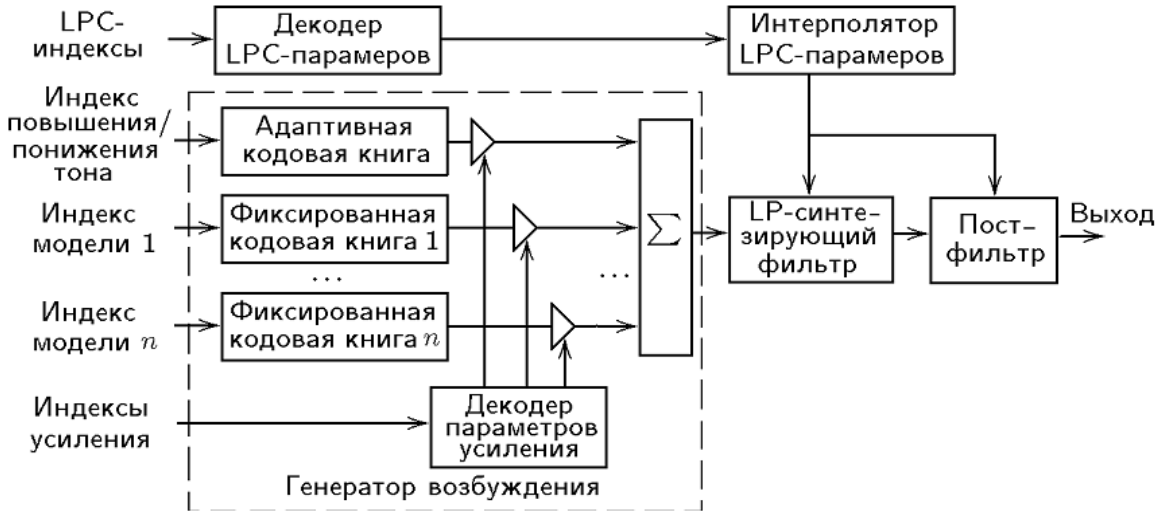


Рис. 2.33. Базовая структура декодера *MPEG*-4 *CELP*

Генератор возбуждения содержит адаптивную кодовую книгу для моделирования периодических компонентов, фиксированные кодовые книги для моделирования случайных компонентов и декодер усиления для восстановления уровня речевого сигнала.

Индексы кодовых книг (повышение/понижение тона для адаптивной кодовой книги, индексы моделей для фиксированных кодовых книг, индексы усиления) используют для генерации возбуждающего сигнала. Сигнал, созданный этим генератором, поступает на вход линейного синтезирующего фильтра с предсказанием (*Linear Predictive Synthesis Filter – LP–Synthesis Filter*). Коэффициенты фильтра восстанавливаются на основе принятых *LPC*-параметров, которые предварительно интерполируются. Значения

этих коэффициентов поступают на вход синтезирующего фильтра. На выходе декодера может быть установлен так называемый «пост-фильтр». Пост-фильтр осуществляет фильтрацию декодированного речевого сигнала с целью улучшения качества восприятия речи. Типичная схема пост-фильтра содержит три основных элемента [2.30]: долговременный пост-фильтр, кратковременный пост-фильтр и устройство масштабирования усиления. Имеются также вспомогательные элементы.

Долговременный пост-фильтр, иногда называемый пост-фильтром основного тона речи, представляет из себя гребенчатый фильтр, спектральные пики которого расположены на частотах, кратных частоте основного сигнала, подлежащего фильтрации. Основная задача кратковременного пост-фильтра заключается в ослаблении частотных составляющих между пиками формант. Устройство масштабирования усиления обеспечивает одинаковый уровень речевого сигнала до и после обработки в пост-фильтре.

Аудиофреймы кодера *MPEG*-4 *CELP*, имеют фиксированную длину. Эти аудиофреймы объединяются в аудиосуперфреймы, длительность которых составляет 400 мс. Применяется неравная защита от ошибок (*UEP*). При этом начало каждого аудиофрейма имеет повышенную защиту от ошибок; биты с нормальной защитой размещаются в оставшейся части данного фрейма. Индексы, указывающие скорость цифрового потока, передаются в канале *SDC* системы *DRM*.

2.7.Процедуры объединения сигналов стереопары в стандартах *MPEG*

Этот режим работы кодеков *MPEG* назван как *Joint Stereo* и применяется при низких скоростях передачи цифровых аудиоданных.

Психоакустические основы

Известно, что частоты, лежащие ниже 150...250 Гц, практически не локализуются слушателем. Во всем остальном спектре звуковых частот они образуют компактные и четкие кажущиеся источники звука. Тем не менее, высокочастотные составляющие стереофонических сигналов, лежащие выше 8000....12000 Гц, также весьма часто практически не влияют на оценку азимута кажущегося источника звука (КИЗ). Это объясняется тем, что энергия звукового сигнала на частотах выше 8000...10000 Гц обычно существенно меньше, чем в области средних частот, где она максимальна для большинства музыкальных инструментов и голосов.

Оценка азимута КИЗ является функцией не только частоты, но и в сильной степени зависит от распределения энергии звукового сигнала по частоте. В оценку азимута КИЗ наибольший вклад оказывают спектраль-

ные составляющие сигнала, энергия которых максимальна. Чем выше энергия спектральной компоненты сигнала, тем в большей степени она определяет оценку азимута КИЗ в пространстве. Кроме того, на частотах выше 1500...2500 Гц оценка азимута КИЗ определяется уже не тонкой временной структурой сигнала, а его огибающей или, точнее говоря, соотношением интенсивностей (энергий) высокочастотных частей спектров сигналов стереопары.

Изложенные выше соображения и лежат в основе процедуры объединения сигналов стереопары. Эта процедура обычно реализуется в ситуации, когда имеющееся в нашем распоряжении количество бит недостаточно для раздельного (независимого) кодирования левого и правого сигналов стереопары. Начиная с определенной частоты, можно вместо левого и правого сигналов стереопары кодировать и передавать их сумму в виде так называемой монофонической добавки. Значение этой частоты зависит от характера распределения энергии по частоте для кодируемой выборки звукового сигнала, оно может меняться от одного аудиофрейма к другому.

В процессе объединения могут появиться заметные на слух искажения как пространственной структуры стереопанорамы, так и тембральные изменения в звучании отдельных музыкальных инструментов и голосов. Для компенсации пространственных искажений, вызванных объединением сигналов стереопары, необходимо дополнительно передать также информацию, достаточную для восстановления после декодирования энергий левого и правого сигналов стереопары в объединенной части спектра. В области же частот выше 8000...10000 Гц в большинстве случаев достаточно для устранения тембральных искажений передать только объединенную часть стереосигнала, то есть так называемую монофоническую добавку, но сохранив при этом общий спектральный баланс для обоих сигналов.

Алгоритмы компрессии стандартов MPEG не содержат четких критериев, определяющих условие перехода кодека в режим объединения сигналов стереопары. Сами же алгоритмы объединения сигналов очень похожи и отличаются только в деталях.

Алгоритм «*Joint-Stereo*» стандарта *ISO/IEC* 14496-3 *AAC*

Процедура кодирования в режиме “*Joint-Stereo*” алгоритма компрессии стандарта *ISO/IEC* 14496-3 *AAC* очень похожа на примененную в *Layer* 3. Поэтому ниже рассматриваются лишь ее основные отличия.

Для каждой субполосы кодирования n путем суммирования квадратов амплитуд коэффициентов МДКП вычисляются энергии левого, правого и суммарного сигналов $e_{n,L}$, $e_{n,R}$, $e_{n,M}$, далее рассчитываются значения координатного множителя ψ_n и амплитуда каждого i–того коэффициента МДКП объединенного сигнала M_i. Коэффициенты МДКП объединенного сигнала передаются вместо соответствующих компонент сигна-

ла левого канала. Соответствующие значения компонент сигнала правого канала приравниваются к нулю ($X_i^{'} = 0$). Далее выполняются стандартные процедуры квантования и кодирования для объединенного сигнала. Заметим, что в отличие от *MPEG Layer* 3 в *AAC* кодируется не сами коэффициенты МДКП, а текущая разность между ними (то есть используется дифференциальная ИКМ). При декодировании объединенных субполосных сигналов левый *L* и правый *R* сигналы стереопары восстанавливаются из одного набора спектральных коэффициентов объединенного субполосного сигнала после его декодирования.

При передаче объединенного сигнала, передаваемого в правом канале, стандартом предписано использование кодовых таблиц Хаффмана типа INTENSITY_HCB и INTENSITY_HCB2, при этом в левом канале применение этих таблиц запрещено. Таблицы INTENSITY_HCB и INTENSITY_HCB2 применяются при кодировании соответственно синфазных и противофазных составляющих объединяемых сигналов стереопары. Информацию о соотношении фаз коэффициентов МДКП исходных сигналов в режиме объединения субполосных сигналов можно получить также посредством флага ms_used_n[1]. Первоначальное соотношение фаз, идентифицированное кодовыми таблицами Хаффмана, меняется из синфазного на противофазное и, наоборот, если соответствующий бит флага *ms_used* установлен для данной субполосы.

При декодировании следует иметь в виду два соображения:

-в системе кодирования AAC координатный множитель кодируется точно так же как и масштабные коэффициенты, т.е. кодами Хаффмана с применением дифференциальных величин с двумя разностными значениями. Если первое значение отсутствует, то дифференциальное декодирование начинается, считая, что последнее значение координатного множителя равно нулю;

-дифференциальное декодирование происходит отдельно для масштабных коэффициентов и координатных множителей. Другими словами, декодер масштабных коэффициентов игнорирует вставленные значения координатных множителей и наоборот. Одни и те же кодовые таблицы используются для кодирования масштабных коэффициентов и координатных множителей. Две функции определяются при декодировании объединенных каналов:

[1] ms_used_n – однобитный флаг который показывает, что данная полоса кодируется с использованием метода *M/S* кодирования [см.6.3, табл.6.10 стандарта ISO/IEC 14496-3]

$$\hbar_n = \begin{cases} +1 & \text{для субполос кодирования правого канала с использованием кодовых таблиц "INTENCSITY\_HCB"} \\ -1 & \text{то же самое, но при использовании кодовых таблиц "INTENCSITY\_HCB2"} \\ 0 & \text{в противном случае} \end{cases},$$

$$\lambda_n = \begin{cases} 1 - 2 \cdot ms\_used_n, & \text{если флаг ms\_mask\_present} = 1 \\ +1 & \text{в противном случае} \end{cases},$$

где флаг ms_used_n принимает значение равное +1 или 0 и двухбитный флаг *ms_mask_present* показывает присутствие маски *MS*.

Декодирование объединенных сигналов происходит следующим образом:

-сигнал левого канала принимается равным объединенному сигналу

$$X_{i,L} = M_i$$

-сигнал правого канала R_n получается путем умножения сигнала левого канала L_n на масштабный коэффициент $scale_n$:

$$X_{i,R} = scale_n \cdot X_{i,L}, \text{ где } scale_n = \hbar_n \cdot \lambda_n \cdot 0,5^{0.25 \cdot \psi_n}.$$

Эффективность процедуры объединения сигналов стереопары

Исследования, выполненные для реальных звуковых сигналов разных жанров с помощью специально разработанной для этой цели экспериментальной установки, показали следующее [2.19]:

-снижение скорости цифрового потока при объединении субполосных составляющих сигналов стереопары сильно зависит от степени корреляции левого и правого сигналов стереопары в субполосах кодирования, от выбранных значений верхних и нижних границ объединяемых субполос и конечно, от структуры самого звукового сигнала (жанра);

-при объединении сигналов стереопары ниже 215 Гц и выше 10465 Гц для длинных блоков и выше 11025 Гц для коротких блоков, среднее значение снижения скорости цифрового потока составляет 2,8% без учета корреляции, а с учетом корреляции – 12,8 % , при установленной скорости цифрового потока 128 кбит/с на канал. При установленной скорости цифрового потока 96 кбит/с на канал, эти значения соответственно равняются 0,4% и 7,2 %. Следовательно, при данном значении скорости цифрового потока доступное для кодирования число бит уже лежит ниже или вблизи требуемого значения;

-результаты экспертных оценок подтверждают, что при установленных скоростях цифрового потока 128 и 96 кбит/с объединение сигналов стереопары на частотах ниже 215 Гц и выше 6847 Гц для длинных блоков и выше 6890 Гц для коротких блоков не приводит к заметным на слух искажениям; дальнейшее увеличение числа объединяемых субполос кодирования дает снижение скорости цифрового потока, но качество восприятия кодированного звукового сигнала при этом ухудшается;

-при скорости цифрового потока 64 кбит/с на канал применение режима объединения сигналов стереопары для большинства стереофонических музыкальных сигналов является не эффективным и приводит к искажению сигнала. Это объясняется, прежде всего, тем, что при данной скорости доступное для кодирования количество бит уже существенно ниже требуемого психоакустической моделью даже при условии объединения ряда субполосных составляющих.

2.8. Учет временной маскировки при кодировании звуковых сигналов

Оценим, опираясь на результаты моделирования, эффективность учета постмаскировки. Для моделирования воспользуемся экспериментальной установкой, изображенной на рис. 2.34. Она реализована в виде программ-

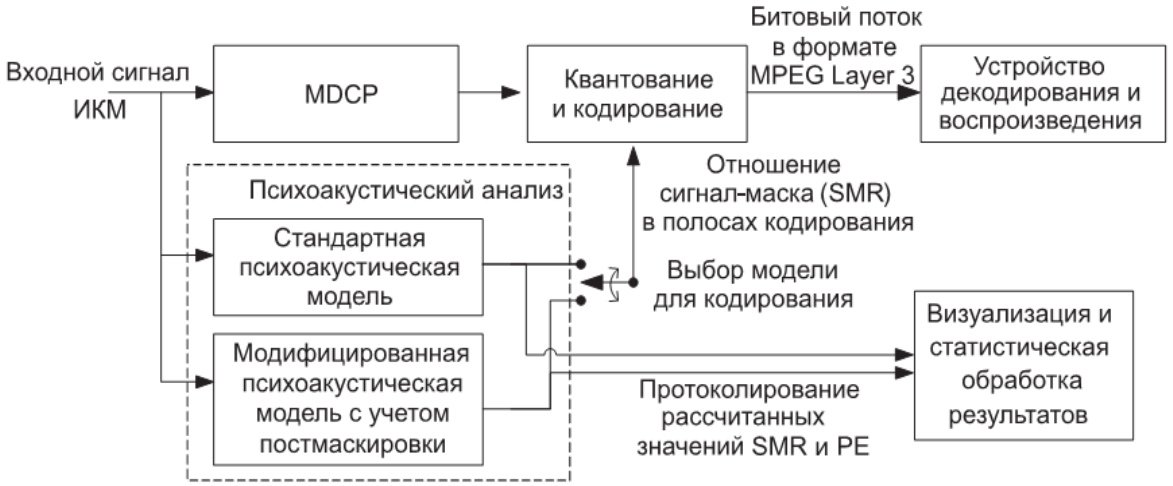


Рис. 2.34. Структурная схема экспериментальной установки

ной модели, написанной на языке *C*. При этом программная модель стандартного кодера *MPEG*-1 *ISO/IEC* 11172-3 *Layer* 3 (*MP*3) дополнена модифицированной психоакустической моделью, а также интерфейсом, необходимым для управления режимами ее работы, а также и для получения нужного массива данных с целью последующего анализа. Кодируемый звуковой сигнал поступает на вход как стандартной (примененной в *Layer* 3), так

и модифицированной психоакустической модели, в которой учет постмаскировки реализован с помощью дополнительного банка цифровых взвешивающих фильтров. Обе модели работают независимо друг от друга. При этом для каждого аудиофрейма вычисляются два отношения сигнал-маска соответственно стандартной и модифицированной психоакустическими моделями. Модификация психоакустической модели 2 кодера стандарта *MPEG* для дополнительного учета влияния постмаскировки представлена на рис. 2.35. Дополнительные по сравнению со стандартной моделью блоки затемнены. Программная модель установки и сами исследования, изложенные ниже, выполнены М.В.Зыряновым [2.20]. Более подробные сведения о моделировании явления временной маскировки приведены в [2.32].

Результат работы любой из этих двух психоакустических моделей может использоваться для квантования и кодирования звукового сигнала. При этом сама процедура квантования и кодирования сигнала, а также временная и частотная сегментации звукового сигнала полностью соответствуют стандарту *MPEG*-1 *ISO/IEC* 11172-3 *Layer* 3.

Программная модель экспериментальной установки позволяет сформировать следующие массивы данных:

-значения отношений сигнал-маска для стандартной и модифицированной психоакустической модели для каждой субполосы кодирования;

-разностные значения отношений сигнал-маска, вычисляемые стандартной и модифицированной моделью, необходимые для расчета экономии бит за счет дополнительного учета постмаскировки.

Массивы перечисленных выше данных записываются в отдельные файлы в *ASCII*–формате, что облегчает контроль и последующий анализ полученных результатов. Для анализа записанных массивов данных программная модель содержит ряд специальных скриптов (небольших программ), написанных на языке *MATLAB* и позволяющих визуализировать полученные результаты. Имеющиеся в *MATLAB* функции построения двумерных и трехмерных графиков, гистограмм распределений, присущая среде *MATLAB* простота программирования и гибкие возможности настройки графического изображения делают этот пакет идеальной платформой для обработки полученных массивов данных.

Зная для каждой из субполос кодирования отношения сигнал-маска, полученные для двух разных психоакустических моделей, легко рассчитать экономию бит за счет дополнительного учета постмаскировки.

Напомним, что изменение числа разрядов в кодовом слове отсчета или коэффициента МДКП на одну единицу приводит к изменению отношения сигнал-шум на 6 дБ. Поэтому число сэкономленных бит, приходящееся на один квантованный отсчет в каждой из субполос кодирования, может быть легко найдено по очень простой формуле, если используется равномерное квантование

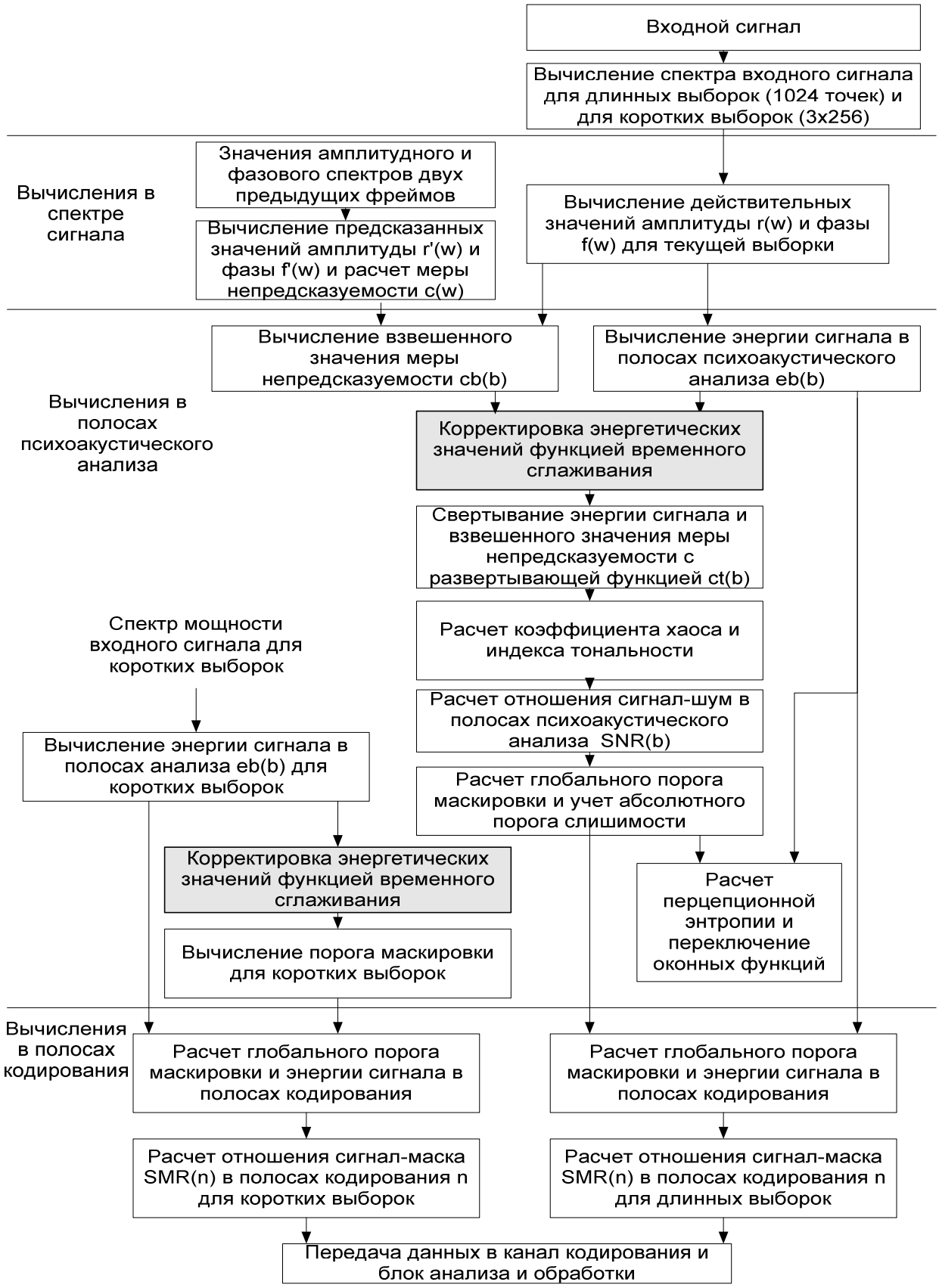


Рис. 2.35. Модифицированная психоакустическая модель 2 стандартов *MPEG*

$$V(b) = bw(b) \cdot \frac{\sum_i SMR(b)(i)}{6,02},$$

где $bw(b)$ – число отсчетов или коэффициентов МДКП в субполосе кодирования b, $SMR(b)(i)$ – разница в значений отношения сигнал-маска в субполосе b для аудиофрейма i. Величина $V(b)$ показывает число сэкономленных бит в субполосе кодирования b для аудиофрейма i. Эти значения вычисляются программной моделью экспериментальной установки, как для секундных интервалов, так и для всего кодируемого звукового отрывка в целом. При этом общее число сэкономленных при учете посмаскировки бит равно для каждого аудиофрейма равно

$V = \sum_b V(b)$,где b – общее число субполос кодирования.

Однако, возможен и другой метод оценки эффективности учета постмаскировки, основанный на расчете значений перцепционной энтропии.

Напомним, что энтропия определена Шенноном как среднее число бит, приходящееся на один символ передаваемых по цифровому каналу данных, при котором не происходит потери информации.

Вычисление энтропии для кодирования переменной x принимающей M возможных состояний производится по формуле:

$$H(x) = \sum_{i=1}^{M} P_i \log_2 \frac{1}{p_i},$$

где p_i – вероятность появления каждого символа. Применительно к кодированию звуковых сигналов, передаваемыми символами являются отсчеты ЗС или коэффициенты МДКП, а вероятность появления каждого символа одинакова и обратно пропорциональна числу уровней квантования. Таким образом, выражение для энтропии приобретает вид

$H = \log_2 M$, когда M равно целой степени по основанию 2,

и

$H = \log_2 M + 1$, когда M не равно целой степени по основанию 2.

Например, для кодирования одного отсчета ЗС или коэффициента МДКП при 65536 возможных уровнях квантования необходимо кодовое слово, содержащее 16 бит.

Понятие перцепционной или психоакустической энтропии в аудиокодировании было введено Джонстоном [2.22]. Оно определяется, как минимальное число бит, необходимое для кодирования одного отсчета звукового сигнала, при котором возникающий шум квантования не воспринимается слухом (лежит по уровню ниже порога слышимости), и декодируемый сигнал при прослушивании не отличается от исходного ЗС. Для этого, мощность шума квантования, определяемая для равномерной ИКМ как

$$P_{Ш.КВ} = \frac{\Delta^2}{12}, \text{ где } \Delta \text{ – шаг квантования,}$$

не должна превышать уровень относительных порогов маскировки, создаваемых кодируемым сигналом.

Таким образом, необходимый для кодирования без потери слухом акустической информации, шаг квантования определяется как

$$\Delta = \sqrt{12THR},$$

где THR - уровень энергии шума кувантования на пороге его маскировки кодируемым сигналом. Величина этого порога определяется при кодировании методом психоакустического анализа кодируемого сигнала. Поскольку, необходимое для представления кодируемой величины $X(b)$ в двоичном виде число бит определяется по формуле:

$$N=\log_2\left(2\cdot \mathrm{nint}\left(\frac{|X|}{\Delta}\right)+1\right),$$

где функция nint вычисляет ближайшее целое, то значение перцепционной энтропии *PE* в битах на отсчет можно вычислить по формуле:

$$PE = \frac{1}{N}\sum_{b=0}^{N-1}\log_2\left(2\cdot \mathrm{nint}\left(\frac{|X(b)|}{\sqrt{12\mathrm{THR}(b)}}\right)+1\right).$$

Здесь $|X(b)|$ – значение энергии кодируемого отсчета (или коэффициента МДКП) в полосе психоакустического анализа b, а $\mathrm{THR}(b)$ – значение порога маскировки в этой полосе.

В психоакустической модели алгоритма компрессии *MPEG Layer* 3, перцепционная энтропия вычисляется на основе отношения вычисленного порога маскировки (thr) к энергии сигнала (eb) по формуле:

$$PE = \sum_b \text{-cbwidth} \cdot \log\left(\frac{\text{thr}+1{,}0}{\text{eb}+1{,}0}\right),$$

где cbwidth – ширина субполосы психоакустического анализа в числе спектральных линиий, thr порог маскировки и eb – уровень энергии сигнала в этой субполосе. Число 1,0 в числителе и знаменателе выражения предотвращают появление нулевых значений в функции под знаком логарифма.

Значения перцепционной энтропии вычисляются в каждой полосе психоакустического анализа, и затем суммируются, умноженные на ширину соответствующей полосы, выраженной в числе спектральных коэффициентов.

Значение перцепционной энтропии, вычисленное по вышеприведенной формуле, не используется в психоакустической модели для оценки минимально достаточного числа бит, необходимого для кодирования сигнала. Оно применяется только для переключения оконных функций (изменения длины выборки кодируемого сигнала), при превышении значением перцепционной энтропии порогового значения. Оно принято равным 1800.

Можно вычислить эквивиалентное пороговому значение перцепционной энтропии, используя формулу, предложенную Джонстоном

$$PE_j = \log_2\left(2\sqrt{\frac{\exp\left(PE_{l3}\right)}{12}} + 1\right)$$

где PE_{l3} – значение перцепционной энтропии, вычисляемое в психоакустической модели алгоритма компрессии *MPEG Layer* 3, а PE_j - значение перцепционной энтропии, вычисляемое по данным Джонстона. Для выполнения этого преобразования, пороговое значение, используемое в стандарте, необходимо нормировать на количество спектральных коэффициентов, используемых психоакустической моделью для вычисления порога маскировки. При частоте дискретизации 44,1 кГц их число равно 465. Выполнив преобразование, получаем, что переключение в режим использования коротких окон при кодировании происходит при превышении необходимой точности квантования в 2,3 бита на отсчет.

На рис. 2.36 представлены значения перцепционной энтропии в битах на отсчет (или коэффициент МДКП) в зависимости от отношения сигнал маска *SMR*, вычисленные двумя рассмотренными выше способами. Здесь же показано пороговое значение, при котором, согласно стандарту,

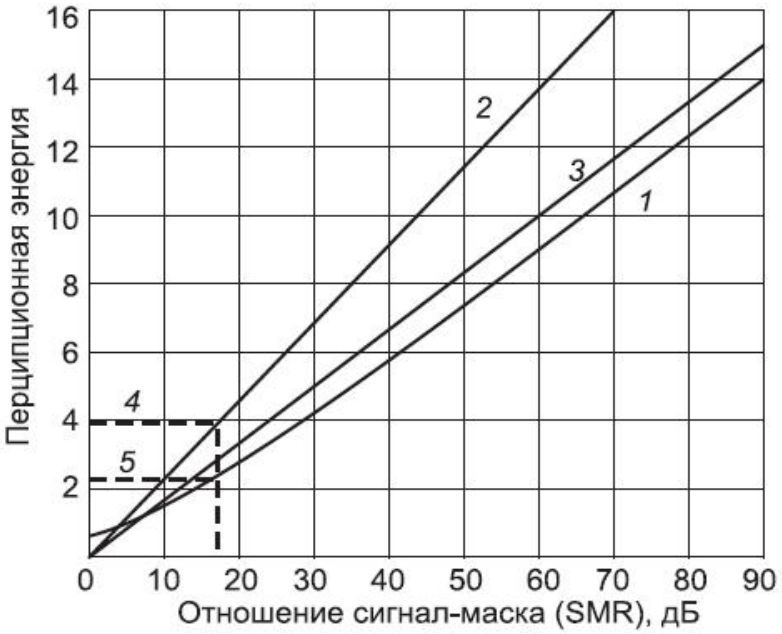


Рис. 2.36. Значения перцепционной энтропии в зависимости от отношения *SMR*, вычисленные по формуле Джонстона и согласно стандарту *MPEG* для психоакустической модели 2

происходит переключение в режим кодирования коротких выборок звукового сигнала.

2.9.Эффективность учета постмаскировки в алгоритмах компресии цифровых аудиоданных

Очевидно, что дополнительная экономия бит за счет учета постмаскировки существенно зависит от динамической структуры звукового сигнала. Для иллюстрации этого вывода ниже представлены зависимости экономии бит, полученные для спокойной с малым динамическим диапазоном (рис. 2.37,*а*) и ритмической с большим значением пик-фактора (рис. 2.37,*б*) музыки. По вертикальной оси для каждого из верхних рисунков отложена величина экономии бит для каждой из субполос кодирования. Это средние значения, отнесенные к одному коэффициенту МДКП в каждой из субполос кодирования. Их общее число в каждой субполосе равно 36. По горизонтальной оси отложены индексы (номера) субполос кодирования. Все они имеют одинаковую ширину равную 750 Гц. Видно, что в обоих случаях получены сходные закономерности: несколько большая (по сравнению со средними частотами 3500…10000 Гц, субполосы кодирования 5…15) экономия бит наблюдается до частоты 3000…3500 Гц (первые пять субполос кодирования) и существенно большая на частотах превышающих 10...11 кГц (субполосы кодирования 15…22). Если на самых низких частотах эта

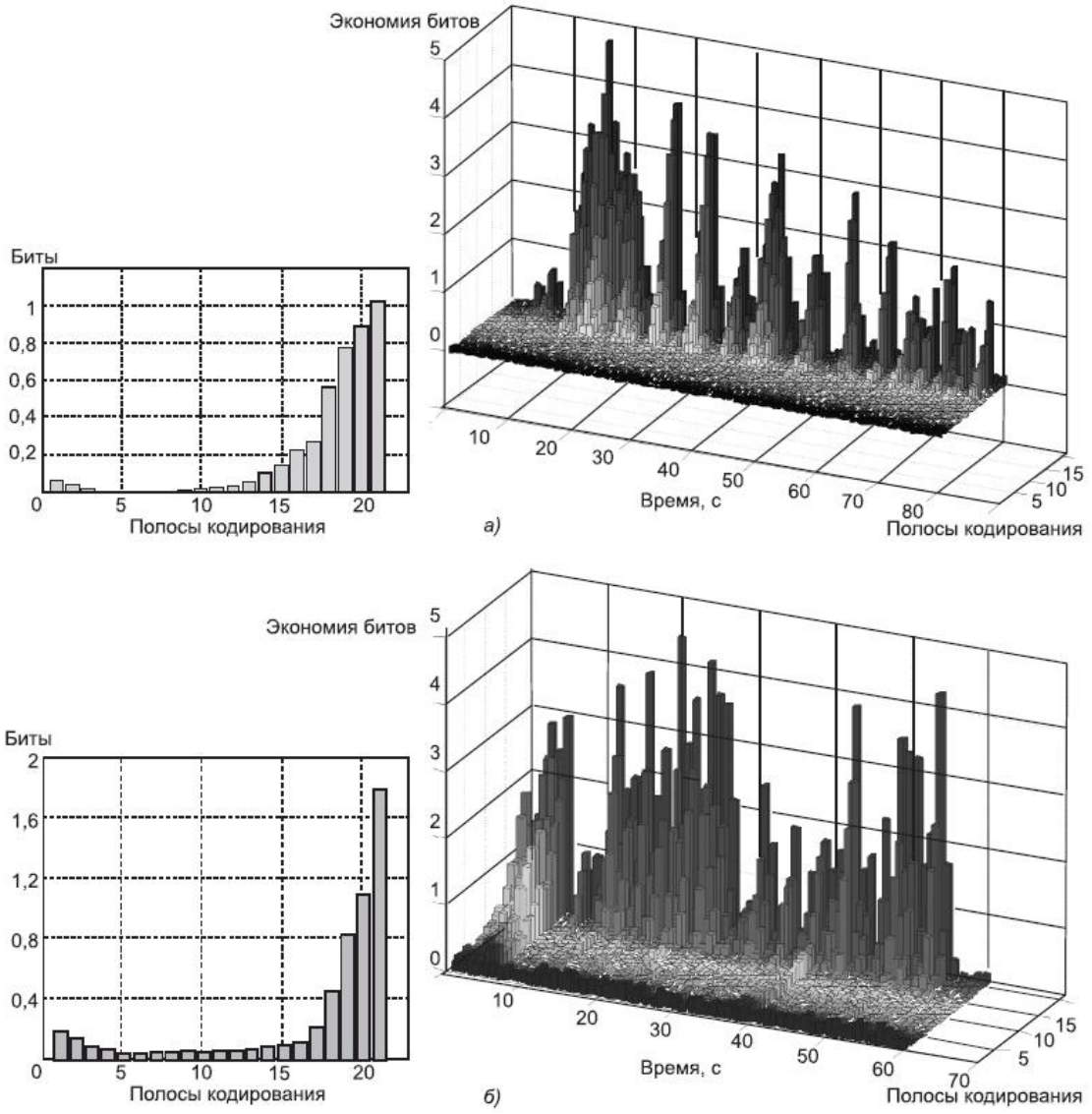


Рис. 2.37. Распределение сэкономленных битов по субполосам кодирования в расчете на один квантованный коэффициент МДКП в каждой из субполос (слева) и трехмерная картина, показывающая изменения числа сэкономленных битов в каждой из субполос кодирования во времени (справа): *а* — для музыкального отрывка с малым динамическим диапазоном и небольшим значением пик-фактора (отрывок классической музыки, спокойная тема); *б* — для музыкального отрывка популярной музыки с большим динамическим диапазоном и большим значением пик-фактора

экономия составляет в среднем доли бит, то на самых высоких частотах она может достигать до 1,5…2 бит при кодировании каждого из коэффициентов МДКП, попадающих в ту или иную субполосу кодирования. В области частот от 3000…3500 до 10000…11000 Гц учет постмаскировки практически не обеспечивает дополнительного выигрыша. Для большей наглядности в нижней части на рис. 2.37,*а* и 2.37,*б* представлены трехмерные графики экономию бит в каждой из субполос кодирования (за счет дополнительного учета постмаскировки) в функции текущего времени. Здесь по одной из горизонтальных осей отложено текущее время, по второй гори-

зонтальной оси – субполосы кодирования в Барк-шкале; по вертикальной оси – сэкономленное число бит.

Имея эти данные (рис. 2.37,*а* и 2.37,*б*) нетрудно подсчитать среднее число сэкономленных бит при кодировании каждой выборки звукового сигнала за счет учета постмаскировки. Напомним, что в *Layer* 3 стандартов *MPEG*-1 *ISO/IEC* 11172-3 или *MPEG*-2 *ISO/IEC* 13818-3 кодируются коэффициенты МДКП. При этом в каждой из субполос кодирования мы имеем по 36 значений коэффициентов МДКП. Используя данные обоих рисунков, можно найти общее число сэкономленных бит при кодировании одной выборки звукового сигнала. Оно равно соответственно около 145 и 345 бит. Напомним, что для прозрачного кодирования выборки звукового сигнала, включающей 1152 отсчета, в среднем необходимо около 2750 бит. Следовательно, учет постмаскировки позволяет в общей сложности в каждом аудиофрейме цифрового потока экономить в среднем от 5,5 (спокойная музыка с малым значением пик-фактора) до 12,5% (ритмическая музыка с большим значением пик-фактора) требуемого для прозрачного кодирования числа бит.

Похожие результаты получаются и для звуковых сигналов других жанров. В табл. 2.1 приведены данные по оценке эффективности учета постмаскировки, усредненные для отрывков звуковых сигналов разных жанров.

Таблица 2.1

Результаты учета влияния постмаскировки на теоретически возможную при кодировании экономию бит, усредненные для различных музыкальных жанров (по данным М.В.Зырянова, [2.20])

Жанры звукового сигнала	Значения энтропии		Выигрыш, бит/отсчет		Выигрыш, %	
	Long (длинная выборка)	Short (короткая выборка)	Long (длинная выборка)	Short (короткая выборка)	Long (длинная выборка)	Short (короткая выборка)
Струнная музыка	1,089	1,011	0,030	0,08	2,740	7,58
Симфоническая	1,148	1,049	0,008	0,04	0,680	4,21
Органная музыка	1,087	0,988	0,006	0,04	0,570	4,07
Электронная музыка	1,139	1,038	0,103	0,16	9,020	15,79
«Металл»	1,109	1,018	0,007	0,03	0,630	2,77
Джаз	1,153	1,053	0,032	0,10	2,790	9,59
Рок-музыка	1,151	1,047	0,029	0,08	2,550	7,74
Поп-музыка	1,129	1,036	0,017	0,05	1,510	4,96
Речь	1,030	0,938	0,168	0,27	16,340	29,17

Интересны также результаты оценки эффективности учета маскировки, полученные австралийскими исследователями (табл.2.2). Эти данные

получены для отрывков звуковых сигналов с явно выраженным ритмом и динамикой, при наличии большого числа выбросов.

Таблица 2.2

Допустимое снижение скорости цифрового потока
при сохранении прозрачного кодирования для высококачественных
звуковых сигналов

Тип звукового сигнала	Минимальное значение скорости цифрового потока при учете только одновременной маскировки, кбит/с	Минимальное значение скорости цифрового потока при учете как одновременной маскировки, так и постмаскировки, кбит/с	Дополнительный выигрыш за счет учета постмаскировки, %
Mariah Carey	133,95	104,95	21,65
Eric Clapton	123,57	98,63	20,17
Susan Veg	93,21	76,31	18,3
Tracy Chapman	110,54	86,06	22,14
Hani Anggraini	104,4	89,21	16,94
Castanets	88,72	71,24	19,71
Jazz	145,36	108,58	25,3
Male Speech	104,24	85,44	18,04

Итак, в алгоритмах компрессии цифровых аудиоданных, основанных на учете феномена маскировки, наибольший эффект дает нам учет одновременной маскировки. Учет постмаскировки может обеспечить дополнительно снижение скорости цифрового потока в пределах от 5..7 % (для сигналов с ровной динамикой) до 18…22% (для сигналов с большой динамикой и значительным количеством выбросов во временной функции). Однако ее учет требует усложнения вычислительных процедур при обработке звуковых сигналов в психоакустической модели кодера. В силу этой последней причины постмаскировка в стандартах *MPEG*-1 и *MPEG*-2 не учитывается. Но появляются экспериментальные образцы кодеков, где это учет уже выполняется.

Дальнейшее снижение скорости цифрового потока при кодировании высококачественных звуковых сигналов радиовещания и телевидения обеспечивают алгоритмы, используемые в стандарте *MPEG*-4 *ISO/IEC* 14496-3. Но в нем, как это было показано ранее, реализованы другие базовые идеи.

2.10. Алгоритм кодирования *MPEG Parametric Stereo*

В данном алгоритме при кодировании дополнительно выделяются, квантуются, а затем кодируются пространственные параметры сигналов стереопары, определяющие структуру пространственной звуковой картины. При этом учитываются особенности слухового восприятия пространственной информации.

Психоакустические основы

Многочисленные психофизические исследования [2.30, 2.31, 2.32] и попытки создания модели бинауральной слуховой системы [2.33, 2.34, 2.35, 2.36, 2.37] дали основание полагать, что слуховой анализатор человека воспринимает пространственные сигналы как функции времени и частоты. Существует весомое доказательство того, что слуховая система анализирует приходящие бинауральные сигналы в частотных полосах (критических полосах слуха, достаточно широких и неодинаковых по ширине полосах частот) без возможности их обрабоки с более высоким частотным разрешением. Разрешение слуховой системы по частоте в этом случае можно описать банком фильтров с шириной полосы пропускания, соответствующей шкале *ERB* (эквивалентной прямоугольной полосе пропускания) [2.38, 2.39, 2.40].

Пространственные характеристики звукового образа меняются достаточно медленно, да и сама слуховая система при их оценке также является инерционной. Время адаптации слуха при оценке азимута источника звука составляет по данным публикаций 30..100 мс, в отдельных работах называют даже цифры 120…150 мс.

При кодировании сигналов стереопары нужно учесть также явление бинауральной демаскировки, повышающей заметность шумов квантования по сравнению с восприятием монофонического сигнала. В разных источниках эта цифра по изменению порога маскировки (при переходе от моно к стереовоспроизведению) колеблется от 3 до 13 дБ в зависимости от степени корреляции сигналов стереопары в субполосах кодирования.

Напомним, что на низких частотах локализацию звуков в основном определяет временное различие (*ITDs*) бинауральной пары сигналов, а на высоких частотах – их разность уровней (*IIDs*). Как известно (раздел 1), эти параметры имеют очень сложную за счет дифракции звуковых волн вокруг головы слушателя зависимость от частоты. Звуковые волны, распространяясь от источнка звука до барабанной перепонки ущей слушателя, искажаются, отражаясь от поверхностей помещения и от ушной раковины, что приводит к сложной частотной зависимости параметров *IIDs* и *ITDs* [2.29]. Кроме того, если несколько источников звука с разными спектральными характеристиками расположены в разных точках пространства, то сигналы,

приходящие на барабанную перепонку, будут обладать даже более сложной частотной зависимостью, потому что они обусловлены интерференцией пространственных сигналов каждого конкретного источника звука.

Исследования особенностей локализации звуков и механизмов оценки их азимута позволили установить, что слуховая система извлекает пространственные параметры звукового образа как функцию времени и частоты. В публикациях при шкалировании слуховых ощущений, связанных с изменением частоты, используют, как известно, несколько шкал: *SPINC*-шкала, Мел-шкала, Барк-шкала, *ERB*-шкала. При разработке алгоритмов компрессии цифровых аудиоданных спектральная разрешающая способность слуха обычно моделируется банком цифровых фильтров со структурой субполос кодирования близкой к *ERB* –шкале.

Хотя слуховая система (вследствие ее инерционности) не может следовать за быстрыми (мгновенными) изменениями пространственных параметров *IIDs* и *ITDs*, тем не менее, это не означает, что слушатели не могут обнаружить эти быстрые их изменения. Медленно меняющиеся *IIDs* и *ITDs* отвечают за изменение положения источника звука в пространстве, в то время, как быстрые их изменения приводят к восприятию так называемой «пространственной диффузности», т.е. к увеличению протяженности воспринимаемых источников звука [2.42].При этом возможность обнаружить наличие или отсутствие *IIDs* и *ITDs* практически не зависит от скорости их изменения [2.43].

Слуховая система обладает вполне определенной конечной разрешающей способностью – слуховые стимулы должны измениться на вполне определенную дискретную величину, чтобы эти изменения были замечены слушателями. Для разности уровней *IIDs* сигналов стерепары равной 0 дБ это пороговое значение мало и составляет около 0,5…1 дБ, для разности уровней *IIDs* = 9 дБ это пороговое значение составляет уже около 1,2 дБ, при разности уровней *IIDs* = 15 дБ – величина порога равна 1,5…2 дБ. Эти данные говорят о слабой зависимости данного параметра от величины разности уровней *IIDs*.

Чувствительность слуха к изменению временного сдвига *ITDs* сигналов стереопары прямо зависит от частоты: на частотах ниже 1000 Гц влияние этого стимула определяющее, а на высоких частотах – практически незаметно. Чувствительность к изменению *ITDs* сильно зависит от частоты. Для частот ниже 1000 Гц эта чувствительность может быть описана чувствительностью к постоянной интераауральной разности фаз (*IPDs*) около 0,05 радиана [2.11, 2.53, 2.59, 2.60]. Исходная величина *ITDs* оказывает некоторое влияние на пороговое значение *ITD*: большое *ITDs* на входе снижает чувствительность к изменениям *ITDs* [2.52, 2.61]. Уровень же воздействия почти не оказывает влияния на чувствительность к *ITD* [2.12]. На более высоких частотах бинауральная слуховая система не в состоянии различить тонкоструктурные (незначительные) временные изменения сигнала. Одна-

ко временные изменения огибающих сигналов стерепары можно обнаружить довольно точно [2.62, 2.63]. Несмотря на такую чувствительность в области высоких частот, локализация источников звука, основанных на оценке *ITDs*, преимущественно осуществляется по низкочастотным сигналам [2.64, 2.65] приблизительно до частоты 1,5…2 кГц.

Чувствительность к изменениям корреляционной связи сигналов сильно зависит от величины их корреляции. При коэффциенте корреляции равном +1 (сигналы когерентны), можно ощутить ее изменения около 0.002, в то время как при корреляции близкой к 0 (сигналы некогерентны) ее изменение должно быть в 100 раз больше [2.66, 2.67, 2.68, 2.69]. Чувствительность к интерауральной когерентности практически не зависит от уровня воздействия до тех пор, пока воздействие значительно превышает абсолютный порог слышимости [2.70]. В области высоких частот корреляция огибающей оказывается подходящим описанием пространственной диффузности [2.47, 2.71]. Чувствительность слуха к изменению корреляции практически не зависит от уровня стимула.

Эти пороги типичны при длительности сигнала не менее 300…400 мс. Если длительность сигнала меньше, пороги обычно повышаются. Например, если длительность существования *IIDs* и *ITDs* во входном воздействии уменьшить с 310 мс до 17 мс, порог может повыситься в 4 раза [2.72]. Чувствительность к интерауральной когерентности также сильно зависит от длительности [2.73, 2.74, 2.75].

При оценке направления слух реагирует на прямой звук, где временные сдвиги не превышают 2 мс. Это явление называют «*законом первого фронта волны*» или «*эффектом Хааса*» [2.76, 2.77, 2.78, 2.79].

Период обновления этих параметров определяется изменением во времени значений пространственных стимулов. Здесь мы имеем дело, как правило, с медленными изменениями.

Три основных пространственных параметра: разность уровней *ILDs*, временной сдвиг *ITDs* и межканальная корреляция или величина взаимной корреляции межканальных сигналов стереопары являются носителями информации о местоположении в пространстве звукового образа. Вследствие бинауральной демаскировки искажения квантования при кодировании стереопары могут быть услышаны. Пороговое различие маскировки между моно и стереослушанием составляет около 3 дБ, более свежие данные дают цифру 6 дБ, а новейшие данные говорят о том, что это различие может достигать 13 дБ. Можно считать, что эти цифры приемлемы и для параметрического кодирования пространственной информации.

Кодирование пространственной информации

Это направление кодирования ЗС получило название *Spatial Audo Coding (SAC*).

Укрупненные структуры кодера (*а*) и декодера (*б*), реализующего этот принцип, представлены на рис. 2.38. В первом блоке кодера выполняется

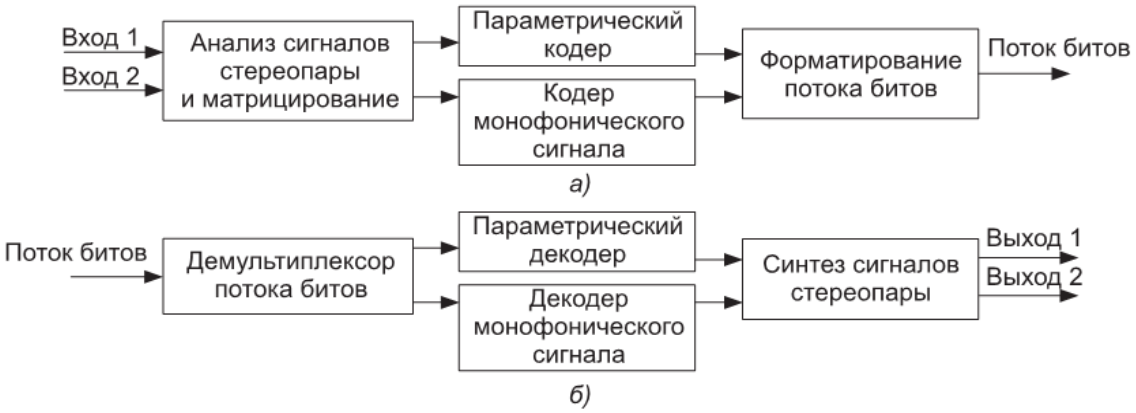


Рис. 2.38. Урупненая структурная схема кодека, реализующего алгоритм *Parametric Stereo*: *а* – кодер; *б* – декодер

анализ сигналов стереопары (*Spatial analysis*) в полосах психоакустического анализа/кодирования, а также объединение исходной пары сигналов путем матрицирования (*Downmix*) в один монофонический сигнал. Выделенные в результате анализа субполосных сигналов пространственные параметры квантуются и кодируются в параметрическом кодере (*Parameter encoder*). Отдельно с использованием обычно кодера стандарта *MPEG*-4 кодируется также монофонический сигнал, полученный путем понижающего матрицирования. После чего оба цифровых потока объединяются, а затем передаются по каналу связи к декодеру. В декодере пространственные параметры выделяются из входного цифрового потока и декодируются. Декодируется также отдельно и монофонический сигнал.

С помощью пространственных параметров из декодированного монофонического сигнала в декодере восстанавливаются исходные левый и правый сигналы стереопары. Так как пространственные параметры подвергаются оценке (в кодере) и используются (в декодере) как функции времени и частоты, поэтому как кодер, так и декодер требуют банк цифровых фильтров, который бы образовал индивидуальную временную/частотную сетку. Частотный анализ на этом этапе должен быть неоднородным (с различным разрешением) из-за особенностей частотного анализа человеческого уха. Временной анализ должен быть медленным (порядка десятков миллисекунд), отражающим концепцию бинауральной инерционности слуха при оценке направления на источники звука за исключением переходных процессов, когда предыдущая информация влияет на временной анализ только в течение нескольких миллисекунд.

Пример более подробной схемы такого кодера представлен на рис. 2.39. Кодер получает сигналы стереопары $x1[n]$, $x2[n]$ с частотой дискретизации

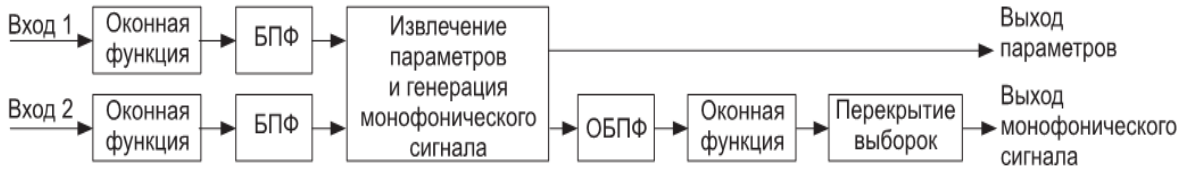


Рис. 2.39. Структурная схема *Spatial*-кодера на основе быстрого преобразования Фурье

fs. Входные сигналы разделяются на сегменты (выборки) с помощью перекрывающихся оконных функций анализа общей длиной *N* с фиксированным числом отсчетов *Nh* в каждом таком сегменте. Если в выборке нет переходного процесса, то длина оконной функции анализа и частота их смены (скорость обновления параметров) должны соответствовать нижней границе постоянной времени, определяющей инерционность бинауральной слуховой системы при локализации источников звука. Эта величина часто принимается равной приблизительно 23 мс. Итак, каждый сегмент ЗС взвешивается с помощью перекрывающихся оконных функций, потом переносится в частотную область с помощью БПФ. В случае переходного процесса используется динамическое переключение окон. Цель динамического переключения двояка: во-первых, нужно учесть эффект Хааса, ивследствие которого только первые 2 мс переходного процесса в реверберационной среде определяют оценку источника звука в азимутальной плоскости; во-вторых, нужно предотвратить появление пред-эха, возникающего при использовании только длинных выборок. Используется 50% перекрытие длинных выборок.

Процедурой переключения окон, показанной на рис. 2.40, управляет детектор переходных процессов.

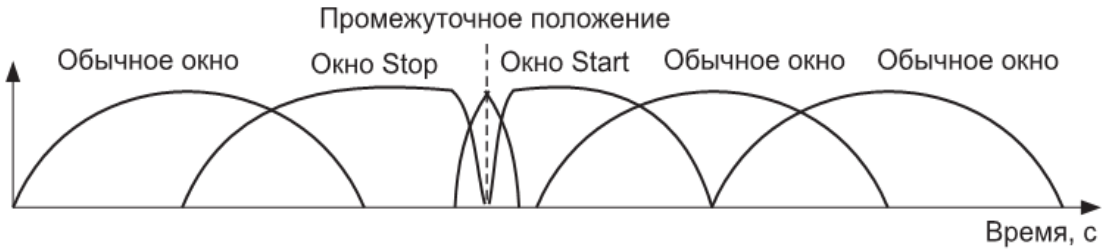


Рис. 2.40. Схема применения оконных функций в *Spatial*-кодере алгоритма *Parametric Audio*

Итак, в процессе кодирования длина выборок (сегментов) может меняться, как это обычно имеет место во всех наиболее эффективных кодеках. Если в пределах длинной выборки переходные процессы не обнаружены, то обрабатывается длинная выборка. Обычно ее длина составляет 23 мс. Каждая выборка взвешивается оконной функцией (*Window*), а затем

преобразуется с помощью быстрого преобразования Фурье (БПФ) в частотную область. Переключение окон (рис. 2.40) выполняется при наличии в выборке переходных процессов (выбросов). При этом учитывается, что только первые 2 мс определяют местоположение источника звука в пространстве. Как только переходной процесс обнаружен, его позиция на временной оси фиксируется коротким окном. Но переход к этому окну выполняется через промежуточную оконную функцию (*Stop window*), переход от короткой выборки к длинной – также через промежуточную оконную функцию (*Start window*).

В следующем блоке после *FFT* формируется монофонический сигнал, а также выделяются, квантуются, кодируются пространственные параметры входных сигналов стереопары, затем они поступают на выход устройства (*Parametr output*). Монофонический сигнал подвергается обратному преобразованию Фурье (*iFFT*), а затем кодируется, после чего поступает на выход устройства (*Mono output*). Для уменьшения искажений, возникающих при выполнении ортогональных преобразований, используется 50% перекрытие выборок.

Длинная выборка содержит 4096 отсчетов при частоте дискретизации 44,1 кГц. В итоге получаем после БПФ 2048 спектральных компонент. Затем спектральные компоненты группируются, образуя субполосы анализа/кодирования в соответствии с *ERB*-шкалой

$$BW = 24{,}7\cdot(0{,}00437\cdot F + 1),$$

где F – центральная частота фильтра субполосного анализа, в Гц. В данном кодере формируется 34 субполосы анализа, номер субполосы b изменяется от 0, 1, 2, ….., B–1. При этом средняя частота для самой низкой субполосы равна 28,7 Гц при $b = 0$, и для самой высокой – соответственно 18,1 кГц при $b = 33$

Для каждой субполосы b анализа/кодирования вычисляются три пространственных параметра:

-разность уровней IID[b] в дБ, как логарифм отношения энергий соответствующих субполосных сигналов левого и правого каналов стереопары

$$\mathrm{IID}[b] = 10\lg\left[\frac{\sum\limits_{k=k_b}^{k_{b+1}-1} X_1[k]X_1^*[k]}{\sum\limits_{k=k_b}^{k_{b+1}-1} X_2[k]X_2^*[k]}\right]$$

-среднее различие по фазе IPD[b], рассчитывается по формуле

$$\mathrm{IPD}[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_2^*[k]\right)$$

-взаимная корреляция IC[b] левого и правого субполосных сигналов

$$\mathrm{IC}[b] = \frac{\left|\sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_2^*[k]\right|}{\sqrt{\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_1^*[k]\right)\left(\sum_{k=k_b}^{k_{b+1}-1} X_2[k]X_2^*[k]\right)}},$$

где $X_1[k]$ и $X_2[k]$ – коэффициенты БПФ субполосных сигналов в полосе анализа b; k_b и k_{b+1} – индексы коэффициентов БПФ, попавших в полосу анализа b; $X_1^*[k]$ и $X_2^*[k]$ – сопряженные значения. Субполосы анализа/кодирования не перекрываются.

Алгоритм обработки сигналов стереопары в матрице *Downmix* имеет вид

$$S[k] = w_1 X_1[k] + w_2 X_2[k],$$

где обычно $w_1 = w_2 = 0{,}5$. Напомним, что для каждого многоканального звукового формата обычно всегда имеется рекомендуемая матрица, с помощью которой он приводится к более низкому формату формату 5.1, этот формат в свою очередь с помощью другой матрицы приводится к формату 2/0, а этот последний – к формату 1/0. Это необходимо для выполнения требования совместимости. Понижающее матрицирование (*Downmix*), выполняемое с помощью фиксированных весовых коэффициентов, несет в себе риск, что энергия (а значит и воспринимаемая громкость) полученного сигнала будет зависеть от взаимной корреляции двух исходных сигналов. Чтобы избежать этого, а также и возможной окраски сигнала и из-за временно – и частотнозависимых взаимных корреляций, весовые коэффициенты $w1$ и $w2$ делают обычно комплексными, чтобы предотвратить возможные компенсации сигналов из-за изменящихся с частотой фазовых соотношений, влияющих на уровень громкости.

После того как получен монофонический сигнал обычно вычисляется также параметр $OPD[b]$. Он определяет среднее в субполосе b различие фазы между входным сигналом стереопары одного из каналов и полученным после матрицирования монофоническим сигналом $S[k]$

$$\mathrm{OPD}[b] = \angle \left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] S^*[k] \right).$$

Декодер для одного из выходных сигналов в данной субполосе *b* берет фазу OPD[*b*], а для сигнала в другом канале в этой же субполосе – фазу OPD[*b*] минус IPD[*b*], ибо параметр IPD[*b*], не указывает, какой из двух сигналов в этой субполосе отстает, а какой опережает *S*[*k*]. С учетом этого параметр OPD вычислен как среднее различие фазы между $X_1[k]$ и *S[k],* где $S^*[k]$ – сопряженное значение по отношению к *S*[*k*]. После этого монофонический сигнал *S*[*k*] переносится во временную область с помощью обратного БПФ. В конце к каждому сегменту применяются окна синтеза и процедура *OLA* (*overlap-add)* – дословно «*сложение с перекрытием*», результатом чего является искомый выходной монофонический сигнал в полной полосе частот. После чего он кодируется обычным способом, например, с помощью алгоритма *AAC+SBR* и передается декодеру.

Квантование пространственных параметров

Монофонический сигнал кодируется/декодируется в соответствии с выбранным для него алгоритмом компрессии (обычно это *AAC+SBR*), далее в блоке *Spatial synthesis* (рис. 2.38,*б*) он используется для реконструкции левого и правого сигналов стереопары с помощью переданных пространственных параметров.

Параметры IID[*b*], IPD[*b*], OPD[*b*] и IC[*b*] квантуются с учетом психоакустики. При кодировании параметра IID[*b*] используется неравномерное квантование т.к. чувствительность слуха к его изменению зависит от его величины. Вектор IID*s* представляет собой возможные значения параметра IID[*b*] с учетом разрешающей способности слуха, обычно это заданный табличным путем набор чисел

$$\mathrm{IIDs} = [\mathrm{IID}_q[0], \mathrm{IID}_q[1], \ldots, \mathrm{IID}_q[30] = [-50, -45, -40, -35, -30, -25, -22, -19, -16, -13, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10, 13, 16, 19, 22, 25, 30, 35, 40, 45, 50].$$

Параметр IID*s* имеет 31 табличное значения. Квантованное значение параметра $\mathrm{IDX}_{\mathrm{IID}}[b]$ для субполосы *b* определяется выражением

$$\mathrm{IDX}_{\mathrm{IID}}[b]] = \arg(\min_i |\mathrm{IID}[b] - \mathrm{IID}_q[i]|)$$

Для параметра IPD_S квантованные значения равны

$$IPD_s = [IPD_q[0], IPD_q[0], ..., IPD_q[7]] = \left[0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{6\pi}{4}, \frac{7\pi}{4}\right].$$

Таких значений всего 8. Этот набор значений также соответствует разрешающей способности слуха к восприятию различий фазы. Напомним, что на низких частотах чувствительность слуха к изменениям временной разности можно описать чувствительностью к постоянной разности фаз.

Квантованное значение параметра $IDX_{IPD}[b]$ для субполосы b вычисляется как

$$IDX_{IPD}[b] = \mathrm{mod}\left[\left(\frac{4IPD[b]}{\pi} + \frac{1}{2}\right), \Lambda_{IPDs}\right],$$

где mod (·) – модуль выражения (·), $\lfloor \cdot \rfloor$- округление в стотрону ближайшего меньшего значения, Λ_{IPDs} – номер элемента в ряду возможных квантованных значений IPDs.

Параметр OPD[b] квантуется также как и IPD[b], его квантованное значение равно

$$IDX_{OPD}[b] = \mathrm{mod}\left[\left(\frac{4OPD[b]}{\pi} + \frac{1}{2}\right), \Lambda_{IPDs}\right].$$

где $IDX_{OPD}[b]$ – квантованное значение параметра OPD для субполосы b.

И наконец, набор квантованных значений для параметра ICs приведен ниже

$$ICs = [IC_q[0], IC_q[0], ..., IC_q[7]] = [1; 0{,}937; 0{,}84118; 0{,}60092; 0{,}36764; 0; -0{,}589; -1], \tag{2.3}$$

и формула для его вычисления имеет вид

$$IDX_{IC}[b]] = \arg(\min_i |IC[b] - IC_q[i]|).$$

Заметим, что последовательность чисел (2.3) основана на заметных разностях корреляции, описанных в [2.69].

Параметры IPD[b] и OPD[b] не передаются для субполос $b > 17$, т. е. начиная с частоты приблизительно равной 2 кГц. В этой области частот вре-

менной сдвиг сигналов стереопары практически не влияет на оценку азимута источника звука.

Таким образом, для каждой выборки передаются 34 значений параметра IID и IC и 17 значений параметров IPD[b] и OPD[b]. При их кодировании используется дифференциальная ИКМ, т.е. кодируется разность текущего значения параметра от его предшествующего значения.

Все пространственные параметры передаются раздельно во времени. В принципе раздельное кодирование индексов Λ (λ = {0, . . . ,Λ - 1}) требует 2Λ - 1 кодовых слов λd = {−Λ + 1, . . . , 0, . . . ,Λ - 1}. Если каждый конкретный индекс λd имеет вероятность появления p(λd), то энтропия $H(p)$ (бит/символ) данного распределения вычисляется как:

$$H(p) = \sum_{\lambda_d=-\Lambda+1}^{\lambda=\Lambda-1} -p(\lambda_d)\log_2(p(\lambda_d)).$$

Учитывая тот факт, что фактическое значение каждого параметра Λ декодеру известно, модуль λmod каждого отдельного индекса λd также может быть закодирован с помощью λmod

$$\lambda_{\text{mod}} = \text{mod}(\lambda_d, \Lambda).$$

Декодер может легко удерживать в памяти передаваемый индекс λ:

$$\lambda[q] = \text{mod}(\lambda_{\text{mod}}[q] + \lambda[q-1], \Lambda),$$

где q – номер текущего фрейма. Энтропия $H(p\text{mod})$ в зависимости от λmod выражается следующим образом:

$$H(p_{\text{mod}}) = \sum_{\lambda_{\text{mod}}=0}^{\Lambda-1} -p_{\text{mod}}(\lambda_{\text{mod}})\log_2(p_{\text{mod}}(\lambda_{\text{mod}})).$$

Из того, что

$$p_{\text{mod}}(0) = p(0),$$
$$p_{\text{mod}}(z) = p(z) + p(z-\Lambda) \quad \text{for } z = \{1, \ldots, \Lambda-1\},$$

следует, что разность энтропии между дифференциальным кодированием и дифференциальным кодированием модуля $H(p) - H(p\text{mod})$ равна

$$H(p) - H(p_{\text{mod}}) = \sum_{\lambda_d=1}^{\lambda_d=\Lambda-1} p(\lambda_d) \log_2 \frac{p(\lambda_d) + p(\lambda_d - \Lambda)}{p(\lambda_d)} + \sum_{\lambda_d=1}^{\lambda_d=\Lambda-1} p(\lambda_d - \Lambda) \log_2 \frac{p(\lambda_d) + p(\lambda_d - \Lambda)}{p(\lambda_d - \Lambda)}.$$

Для неотрицательных вероятностей $p(\cdot)$ следует, что

$$H(p) - H(p_{\text{mod}}) \geq 0.$$

Дифференциальное кодирование модуля приводит к энтропии, меньшей, чем при не модульном дифференциальном кодировании. Однако при этом выигрыш в скорости цифрового потока относительно мал: около 15% для параметров IPD и OPD и практически отсутствует для параметров IID и IC.

Величина энтропии на символ при дифференциальном кодировании модуля и итоговый вклад в конечную скорость цифрового потока представлены в табл. 2.3. Эти числа были получены путем анализа 80 разных отрывков реальных звуковых сигналов. Суммарная скорость передачи этих параметров равна 7,7 кбит/с. Для 20 субполос кодирования получим уже значение скорости 4,5 кбит/с. Период обновления этих данных равен 23 мс. Это длина выборки звукового сигнала.

Таблица 2.3

Возможные значения скорости цифровых потоков при передаче пространственных параметров сигналов стереопары

Parameter	Bits/symbol	Symbols/s	Bit rate (bps)
IID	1.94	1464	2840
IPD	1.58	732	1157
OPD	1.31	732	959
IC	1.88	1464	2752
Total	—	—	7708

Отсутствие передачи параметров IPD и OPD в субполосах ниже 10-й сопровождается существенным ухудшением качества при прослушивании. При 34-х субполосах кодирования и длительности аудиофрейма равной 23 мс скорость передачи пространственных параметров составляет примерно

8 кбит/с. Ее значение равно около 1,5 кбит/с при 20 субполосах анализа и длине фрейма 46 мс, если нет передачи параметров IPD и OPD. Если передавать два параметра IID и IC в каждой субполосе анализа/кодирования, то может быть получена экономия скорости около 27% по сравнению с полной передачей (табл.2.3) всех параметров.

Декодирование пространственных параметров

В декодере нужно создать два выходных сигнала $y1[n]$ и $y2[n]$. Они должны обладать переданными пространственными параметрами. Заметим, что декодер, построенный на основе БПФ (рис. 2.41), дает дискретный

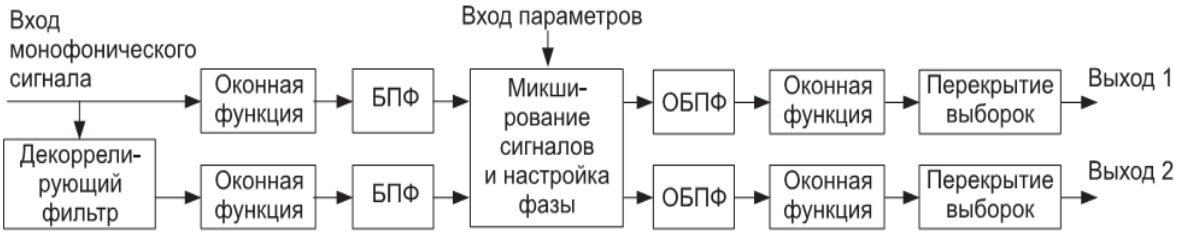


Рис. 2.41. Структурная схема *Spatial*-декодера на основе БПФ

спектр с постоянным шагом по частоте, зависящем от величины частоты дискретизации и длины выборки (сегмента, для которого выполняется БПФ). При выборе требуемой разрешающей способности на низких частотах получаем в этом случае излишнюю разрешающую способность на верхних частотах. Чтобы создать два выходных сигнала с переменной (т.е. зависящей от параметра) когерентностью, у второго сигнала должна быть такая же временная и спектральная огибающая, как и у входного монофонического сигнала, но он должен быть некогерентным с точки зрения быстроменяющейся, тонкоструктурной формы волны. Такой некогерентный (или ортогональный) сигнал $sd[n]$ получается с помощью операции свертки входного монофонического сигнала $s[n]$ с импульсной характеристикой всепропускающего декоррелирующего фильтра $hd[n]$. Наиболее простой всепропускающий декоррелирующий фильтр получается на основе обычной линии задержки. Этот метод хорошо работает при условии, что величина задержки достаточно большая, чтобы обеспечить несколько вершин и впадин в каждом фильтре. Т.к. ширина полосы пропускания фильтра больше в области высоких частот, хотелось бы, чтобы задержка зависела от частоты, уменьшаясь в области ВЧ. Дополнительное преимущество частотно-зависимой задержки в том, что она не влияет на гармонические эффекты на выходе гребенчатых фильтров.

Декоррелированный $S_d[n]$ сигнал (рис. 2.41) получается путем свертки монофонического сигнала с импульсной характеристикой всепропускающего декоррелирующего фильтра вида

$$h_d[n] = \frac{2}{N_s} \sum_{k=0}^{N_s/2} \cos\left[\frac{2\pi kn}{N_s} + \frac{2\pi k(k-1)}{N_s}\right],$$

где $0 \le n \le N_s - 1$, а $N_s = 640$.

Рекомендуется использование для этой цели фильтра Шредера [2.88].

Затем выполняются такие же, как в кодере, операции сегментации, взвешивания с помощью оконной функции и преобразования, чтобы получить представление в частотной области $S[k]$ и $Sd[k]$ для входного моносигнала $s[n]$ и его декоррелированного варианта $sd[n]$. Следующий этап состоит из вычисления линейных комбинаций двух входных сигналов, чтобы попасть в два выходных сигнала в частотной области $Y1[k]$ и $Y2[k]$.

Выходные сигналы получаются с помощью выражения вида

$$\begin{bmatrix} Y_1[k] \\ Y_2[k] \end{bmatrix} = R_b \begin{bmatrix} S[k] \\ S_d[k] \end{bmatrix}$$

как динамический процесс умножения каждого из них в субполосе анализа b на множитель

$$R_b = \sqrt{2}\Pi[b]]\mathsf{A}[b]\mathsf{V}[b].$$

Для проведения этих вычислений используются три матрицы:

$$\Pi[b] = \begin{bmatrix} e^{j\mathrm{OPD}[b]} & 0 \\ 0 & e^{j\mathrm{OPD}[b]-j\mathrm{IPD}[b]} \end{bmatrix}; \quad \mathsf{A}[b] = \begin{bmatrix} \cos(\alpha[b]) & -\sin(\alpha[b]) \\ \sin(\alpha[b]) & \cos(\alpha[b]) \end{bmatrix};$$

$$\mathsf{V}[b] = \begin{bmatrix} \cos(\gamma[b]) & 0 \\ 0 & \sin(\gamma[b]) \end{bmatrix},$$

где

$$\alpha[b] = \begin{cases} \pi/4 & \text{для } (\mathrm{IC}[b], c[b]) = (0,1); \\ \mathrm{mod}\left[\frac{1}{2}\,\mathrm{arctg}\left(\frac{2c[b]\mathrm{IC}[b]}{c[b]^2-1}\right), \frac{\pi}{2}\right] & \text{в других случаях;} \end{cases}$$

$$\gamma[b] = \mathrm{arctg}\sqrt{\frac{1-\sqrt{\mu[b]}}{1+\sqrt{\mu[b]}}}; \quad \mu[b] = 1 + \frac{4\mathrm{IC}^2[b]-4}{(c[b]+1/c[b])^2}; \quad c[b] = 10^{\mathrm{IID}[b]/20}.$$

Матрица **A** содержит данные об изменении фазы; решения матриц **P** и **V** приведены ниже. При выполнении этой процедуры должны выполняться следующие требования:

-соотношение энергий выходных сигналов в каждой субполосе должно соответствовать параметру IID;

-корреляция сигналов в каждой субполосе должна соответствовать параметру IC;

-средняя энергия двух выходных сигналов должна равняться средней энергии входного монофонического сигнала;

-среднее различие фазы выходных сигналов в каждой субполосе должно соответствовать параметру IPD;

-среднее различие фазы между сигналами $S[k]$ и $Y1[k]$ должно быть равным значению OPD.

Если параметры IPD/OPD не передаются, параметры IC могут стать отрицательными. Этот случай требует другой матрицы **R.** Подходящее решение получается, если мы увеличим $S[k]$ в суммарном выходном сигнале (т.е. $Y1[k] + Y2[k]$). Это приведет к изменению матрицы $\mathbf{R}_A[b]$:

$$\mathbf{R}_A[b] = \begin{bmatrix} c_1 \cos(\nu[b] + \mu[b]) & c_1 \sin(\nu[b] + \mu[b]) \\ c_2 \cos(\nu[b] - \mu[b]) & c_2 \sin(\nu[b] - \mu[b]) \end{bmatrix},$$

где

$$c_1[b] = \sqrt{\frac{2c^2[b]}{1 + c^2[b]}},$$

$$c_2[b] = \sqrt{\frac{2}{1 + c^2[b]}},$$

$$\mu[b] = \frac{1}{2} \arccos(\mathrm{IC}[b]),$$

$$\nu[b] = \frac{\mu[b](c_2[b] - c_1[b])}{\sqrt{2}}.$$

Итак, субполосные выборки ЗС преобразуются во временную область, взвешиваются с помощью оконной функции (используются такие же окна, как в кодере) и объединяются с помощью процедуры *OLA*.

Применение QMF-фильтров при формировании субполос анализа и синтеза

Декодер на основе БПФ, как сказано выше, требует достаточно большой длины выборки при выполнении БПФ, чтобы обеспечить значительное разрешение по частоте в области низких частот. В результате раз-

решение на высоких частотах оказывается выше, чем нужно, и, следовательно, повышаются требования к памяти декодера на основе БПФ. Уменьшить разрешение на высоких частотах звукового диапазона при сохранении требуемого значения на низких частотах можно, используя квадратурные зеркальные фильтры *QMF*. Если быть более точным, то используемые в данном случае гибридные комплексно-модулированные квадратурные зеркальные фильтры являются расширением банка фильтров, применяемого в алгоритме *SBR* [2.5, 2.6, 2.90]. Обобщенная структура параметрического стереодекодера на базе *QMF*-фильтров показана на рис. 2.42.

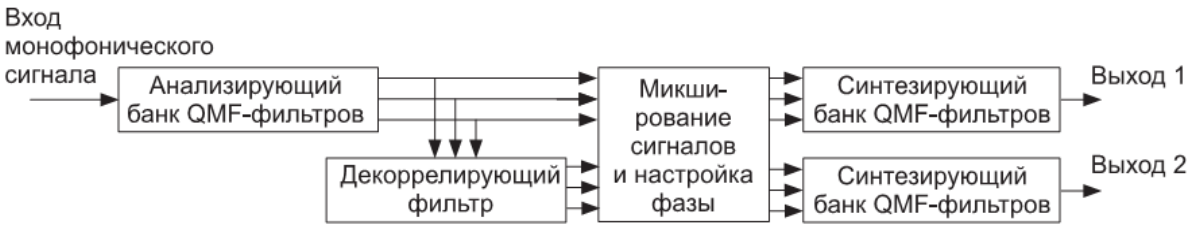


Рис.2.42. Анализирующий и синтезирующий банки фильтров *Spatial*-декодера для получения стереофонического сигнала из монофонического сигнала

Здесь показан пример синтеза стереофонического сигнала из монофонического с помощью анализирующего банка гибридных квадратурных фильтров (*Hybrid QMF analysis*). С их помощью спектр входного сигнала разделяется на полосы анализа с учетом критических полос слуха и его разрешающей способности по частоте. Затем каждый из этих субполосных сигналов поступает на соответствующий декоррелирующий фильтр (*Dekorr. filter*). Затем сигналы с выхода банка фильтров и некоррелированная версия этих сигналов подаются на устройство микширования (матрицу) и регулировки фазы (*Mixing and phase adjustment*). Это устройство создает два выходных сигнала в *QMF*-области с пространственными параметрами, соответствующими переданным. В конце выходные сигналы подаются на банк синтезирующих фильтров, чтобы получить конечные выходные сигналы. Банк анализирующих фильтров состоит из каскадного соединения двух банков фильтров. Его структура показана на рис. 2.43. Первый банк фильтров аналогичен банку фильтров, используемому в алгоритме *SBR*. В конце два банка гибридных *QMF* (*Hybrid QMF synthesis*) фильтра генерируют два выходных сигнала.

Субполосные сигналы, создаваемые банком фильтров, получаются путем выполнения операции свертки входного сигнала с импульсными характеристиками $h_k[n]$ анализирующих фильтров вида

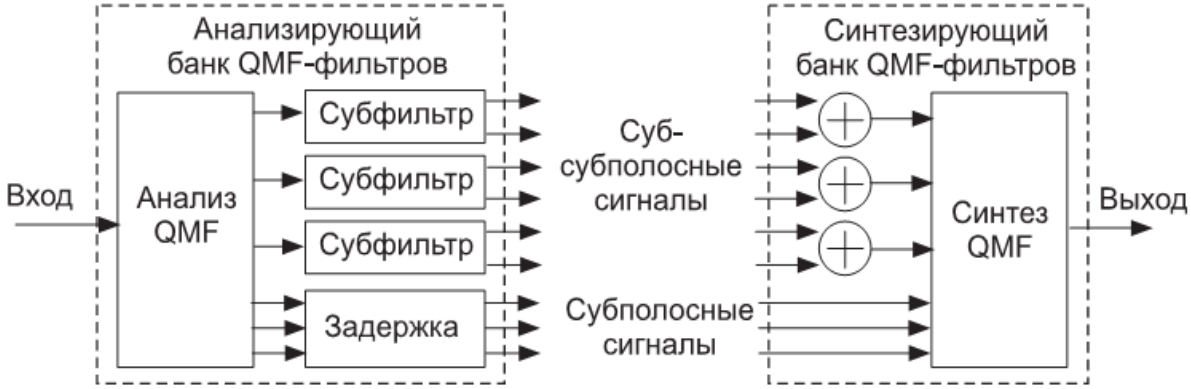


Рис. 2.43.Структуры анализирующего и синтезирующего банков *QMF*-фильтров

$$h_k[n] = p_0[n] \exp\left\{ j\frac{\pi}{4K}(2k+1)(2n-1) \right.$$

где $p_0[n]$, при $n = 0, \ldots, Nq-1$, окно-прототип фильтра, $K = 64$ – число выходных каналов, k – номер субполосы ($k = 0, \ldots, K-1$), а $Nq = 640$ – порядок фильтра. На выходе фильтров, частоту дискретизации понижают в K раз.

Амплитудно-частотные характеристики первых четырех субполос ($k = 0, \ldots, 3$) банка анализирующих QMF-фильтров показаны на рис. 2.44.

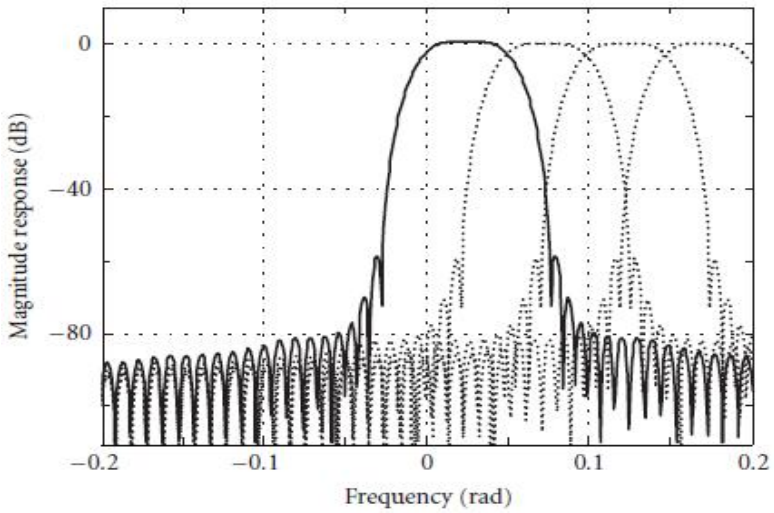


Рис. 2.44. Амплитудно-частотные характеристики первых четырех субполос банка 64-полосных анализирующих фильтров. АЧХ при k = 0 выделена сплошной линией

Субполосные сигналы с пониженной частотой дискретизации $Sk[q]$, имеющие нижние значения субполос, подаются на второй банк комплексно-модулированных фильтров (банк субфильтров) для дальнейшего повышения разрешения по частоте. Сигналы остальных полос задерживаются,

чтобы скомпенсировать задержку, возникающую в банке субфильтров. Выход банка гибридных (т.е. совмещенных) фильтров обозначается $Sk,m[q]$, где k – номер субполосы исходного банка *QMF*-фильтров, а m – номер фильтра в банке субфильтров.

Банк субфильтров содержит фильтры порядка $Ns = 12$ с импульсными характеристиками

$$G_{k,m}[q] = g_k[q] \exp\left\{ j\frac{2\pi}{M_k}\left(m + \frac{1}{2}\right)\left(q - \frac{N_s}{2}\right)\right\},$$

где $g_к[q]$ – окно-прототип, связанный с банком *QMF-k, q* – номер отсчета, а M_k – количество суб-субполос в субполосе k $(m = 0, \ldots, M_k - 1)$.

В табл.2.4 приведены номера суб-субполос M_k как функции от банка фильтров k для 34 и 20 полос анализа. В качестве примера, на рис. 2.45 приведены амплитудно-частотные характеристики четырехполосного банка

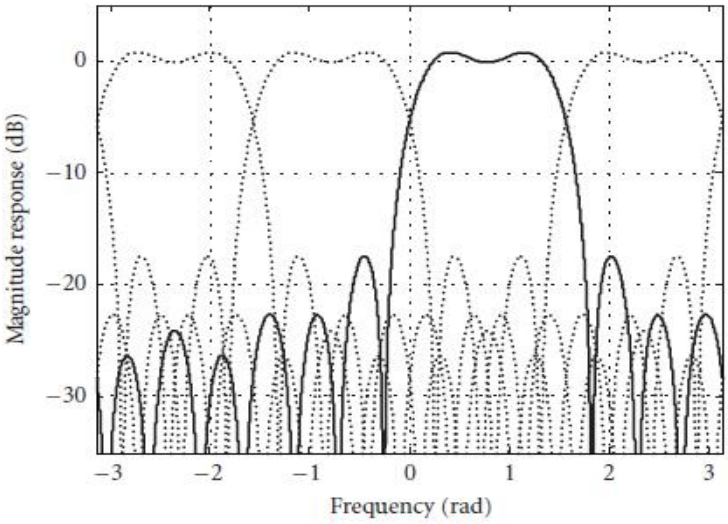


Рис. 2.46. Амплитудно-частотная характеристика четырехполосного банка субфильтров. АЧХ при m = 0 выделена сплошной линией

субфильтров ($M_k = 4$). Очевидно, что из-за ограниченной длины прототипа ($N_S = 12$), ослабление в полосе ослабления только порядка 20 дБ.

Таблица 2.4

Значения M_k для первых пяти субполос QMF

QMF subband (k)	M_k ($B = 34$)	M_k ($B = 20$)
0	12	8
1	8	4
2	4	4
3	4	1
4	4	1

Результатом такой структуры является 91 (при B = 34) либо 77 (B = 20) выходных сигналов $S_{k,m}[q]$ с пониженной частотой дискретизации, а также их декоррелированные аналоги $S_{k,m,d}[q]$, доступные для дальнейшей обработки.

В общей сложности банк *QMF*-фильтров разделяет спектр звукового сигнала либо на 34, либо на 20 разных по ширине субполос анализа и кодирования. Для каждой субполосы в кодере выделяются, квантуются, кодируются параметры IID[b], IPD[b], OPD[b], IC[b]. При этом кодовое слово каждого параметра содержит 4 бита.

Декоррелирующий фильтр можно реализовать разными способами. Элегантный метод представлен в [2.24]; он включает в себя ревербератор. Альтернативный менее сложный способ включает в себя (частотно - зависимую) задержку T_k, время задержки которой зависит от номера субполосы k *QMF*-фльтра.

Следующим этапом пространственного синтеза на основе *QMF*-фильтров является процедуры смешивания и регулировки фазы. Для каждой суб-субполосной сигнальной пары $S_{k,m}[q]$, $S_{k,m,d}[q]$ генерируется выходная сигнальная пара $Y_{k,m,1}[q]$, $Y_{k,m,2}[q]$:

$$\begin{bmatrix} Y_{k,m,1}[q] \\ Y_{k,m,2}[q] \end{bmatrix} = \mathbf{R}_{k,m} \begin{bmatrix} S_{k,m}[q] \\ S_{k,m,d}[q] \end{bmatrix}.$$

Матрица смешивания $\mathbf{R}_{k,m}$ определяется следующим образом. В каждой группе параметров IID, IPD, OPD, IC для каждого параметра субполоса b занимает определенную полосу частот в определенный момент времени. Частотный диапазон зависит от особенностей анализа полос кодером (т.е. от группировки отсчетов БПФ), тогда как момент времени зависит от того, как кодер выполнит сегментацию во временной области. Если кодер спроектирован верно, то частотно-временное положение каждой группы параметров совпадает с определенным номером суб-субполосы или нескольких суб-субполос в области *QMF*. Для некоторых номеров отсчетов матрицы смешивания точно такие же, как их аналоги, основанные на БПФ. Для отсчетов, находящихся посередине, матрицы смешивания линейно интерполируются (т.е. их вещественные и мнимые части интерполируются отдельно). После смешивания следует пара банков гибридных синтезирующих фильтров (один на каждый канал). Процедура синтеза тоже состоит из двух этапов. На первом этапе выполняется суммирование суб-субполос m, относящихся к одной субполосе k:

$$Y_{k,1}[q] = \sum_{m=0}^{M_k-1} Y_{k,m,1}[q],$$

$$Y_{k,2}[q] = \sum_{m=0}^{M_k-1} Y_{k,m,2}[q].$$

И наконец, для получения выходного стереосигнала выполняется повышение частоты дискретизации и свертка с импульсными характеристиками синтезирующих фильтров, которые аналогичны анализирующим фильтрам. Тот факт, что для алгоритмов *PS* и *SBR* используются банки фильтров одинаковой структуры, позволяет легко и без особых затрат совместить *SBR* и параметрическое кодирование в одном декодере, [2.23, 2.24, 2.91, 2.92]. Эта комбинация известна как *AAC Plus* и рассматривается как стандарт для *MPEG*-4 в качестве кодека *HE-AAC/*PS [2.93].

Структура декодера HE-AAC/PS показана на рис. 2.46. Входящий

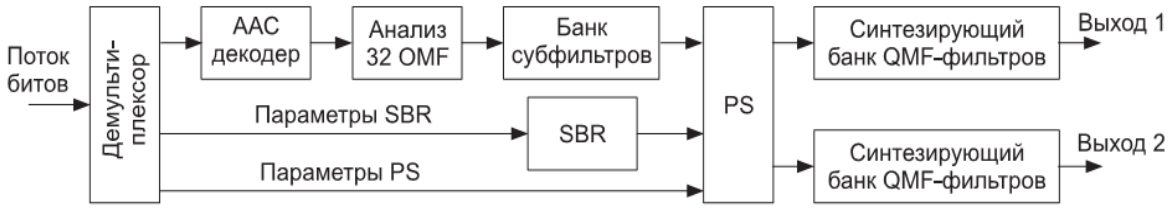


Рис. 2.46. Структурная схема декодера с дополнительной передачей *PS*-параметров

цифровой поток демультиплексируется в поток *AAC* с ограниченной шириной полосы, параметрами *SBR* и данными о параметрическом кодировании пространственной информации. Поток данных *AAC* декодируется с помощью *AAC*-декодера и подается на банк 32-полосных анализирующих фильтров. Результирующий широкополосный монофонический сигнал преобразуется в стереосигнал с помощью уже изложенной выше процедуры *PS*. В конце два банка гибридных синтезирующих фильтров создают выходной сигнал. Более подробную информацию об алгоритме *AAC Plus* можно найти в [2.23, 2.92]

Результаты прослушивания

Экспертизы по оценке качества алгоритма PS преследовали две цели:

-во-первых, необходимо было оценить максимально возможное качество звучания, которое можно получить, используя лежащую в основе кодера пространственную модель. Достаточно большое число авторов считали,

что параметрическое кодирование пространственной информации стереосигнала эффективно только на низких скоростях, так как при нем невозможно в принципе добиться прозрачного кодирования [2.20, 2.21, 2.22], когда искажения, вызванные компрессией цифровых аудиоданных, незаметны на слух;

во-вторых, нужно было найти максимальную общую скорость передачи цифровых аудиоданных, при которой параметрическое кодирование пространственной информации еще будет давать выигрыш по сравнению со стандартными методами кодирования, когда левый и правый сигналы кодируются отдельными устройствами.

- и последний опыт производился, чтобы установить действительную выгоду параметрического кодирования в целом кодере. Для этой цели производилось сравнение между современным стереокодером (*AAC Plus*) и таким же кодером, дополненным алгоритмом «параметрического стерео», (*AAC Plus/PS*, рис. 2.46).

Контрольное прослушивание 1. В этом эксперименте участвовали 9 слушателей. У каждого из них уже был опыт оценки качества аудиокодеков с компрессией цифровых аудиоданных. Кроме того они были специально

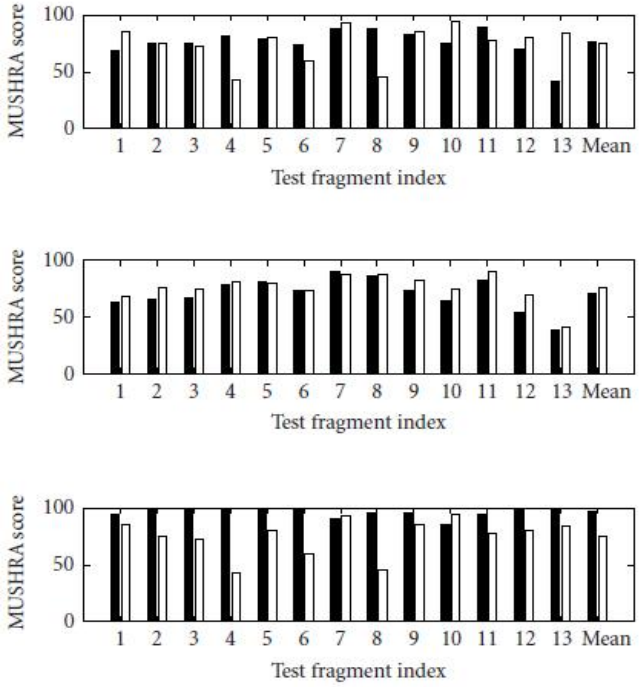


Рис. 2.47. Качество кодированного сигнала: средние оценки экспертов как функция от номера фрагмента и различной конфигурации кодека. Верхний график показывает результат оценки для параметрического стерео при скорости передачи пространственных параметров 8 кбит/с (черные столбики) и кодека *MP*3 со скоростью 128 кбит/с (белые столбики). На среднем графике приведены результаты для параметрического стерео при скорости передачи пространственных парамеров 5 кбит/с (черные столбики) и 8 кбит/с (белые столбики). Самый нижний график показывает результаты оценки эталона (черные столбики) и кодека *MP*3 со скоростью 128 кбит/с (белые столбики)

проинструктированы, чтобы оценивать как качество самой стереопанорамы (протяженности КИЗ, их место в пространстве), так и другие заметные артефакты компрессии. В ходе эксперимента, проводимого по методу *MUSHRA*, слушателям нужно было сравнить качество восприятия нескольких фрагментов реальных ЗС, прошедших соотвествующие кодеки, с оригиналом (необработанный сигнал) по 100-балльной шкале. Отрывки реальных сигналов были взяты с компакт-дисков. Воспроизведение выплнялось с помощью головных телефонов *Stax Lambda Pro*.

Обработанные (прошедшие кодек) фрагменты ЗС были получены с помощью кодеков:

-*MPEG*-1 *Layer* 3 (*MP*3) при установленной скорости цифрового потока равной 128 кбит/с и с максимально возможными параметрами качества;

-на основе БПФ без использования кодера монофонического сигнала, как было сказано выше (т.е. предполагается прозрачное кодирование монофонического сигнала и скорости передачи пространственых параметров равной 8 кбит/с;

-на основе БПФ без кодека монофонического сигнала со скоростьюпередачи пространственных параметров равной 5 кбит/с использованием 20 частотных полос анализа/кодирования вместо 34;

-качество эталонного сигнала также подвергалось субъективной оценке, его предъявление было не известно слушателям.

Все отобранные отрывки (табл.2.5) являются стереофоническими, Это ИКМ-сигналы с разрешением 16 бит/отсчет и с частотой дискретизации 44,1 кГц. Каждый отрывок можно слушать столько раз, сколько хочется, и можно в режиме реального времени переходить к одной из четырех версий фрагмента. Длительность каждого фрагменты составляла около 10 секунд, принадлежали они к разным музыкальным жанрам и были взяты с компакт-дисков.

Таблица 2.5

Перечень отобранных отрывков реальных звуковых сигналов

Item index	Name	Origin/artist
1	Starship Trooper	Yes
2	Day tripper	The Beatles
3	Eye in the sky	Alan Parsons
4	Harpsichord	MPEG si01
5	Castanets	MPEG si02
6	Pitch pipe	MPEG si03
7	Glockenspiel	MPEG sm02
8	Plucked string	MPEG sm03
9	Yours is no disgrace	Yes
10	Man in the long black coat	Bob Dylan
11	Vogue	Madonna
12	Applause	SQAM disk
13	Two voices	Left = MPEG es03 = English female Right = MPEG es02 = German male

Выбранные фрагменты реальных ЗС оказались самыми критичными с точки зрения заметноси искажений.

Средние оценки всех экпертов представлены на рис. 2.47. Сверху показано среднее значение оценок *MUSHRA* при скорости передачи протранственных параметров равной 8 кбит/с (черные столбики) и кодека *MP*З со скоростью передачи данных 128 кбит/с (белые столбики) как функция от номера фрагмента ЗС. Столбики справа показывают среднее арифметическое всех исследуемых отрывков. Большинство фрагментов получили примерно одинаковые оценки, кроме фрагментов 4, 8, 10 и 13. Фрагменты 4 («*Harpsichord*») и 8 («*Plucked string*») значительно более качественные при использовании параметрического стереокодирования. В них много тональных компонент, что является проблемой при кодировании формы сигнала из-за высокого уровня шума квантования в этом случае. С другой стороны, фрагмент 10 («Человек в длинном черном пальто») и 13 («Два голоса») получили более высокие оценки при *MP*З. Фрагмент 13 представляет собой (искусственно) большую независимость каналов, что практически потерялось после параметрического стереодекодирования. В итоге оба кодера получили равные оценки.

На среднем рисунке показаны результаты для параметрического стерео со скоростью 5 кбит/с (черные столбики) и 8 кбит/с (белые столбики). В большинстве случаев кодер со скоростью передачи пространтсвенных параметров 8 кбит/с обладает лучшим качеством, чем при скорости 5 кбит/с, кроме фрагмента 5 («Кастаньеты») и 7 («*Glockenspiel*»). В итоге качество кодера при скорости передачи пространтственных параметров 5 кбит/с незначительно ниже, чем при скорости 8 кбит/с, что показывает медленно убывающее отношение скорость/качество при параметрическом кодировании.

Самый нижний график показывает результаты для *MP*З со скоростью передачи 128 кбит/с (белые столбики) и скрытого эталона (черные столбики). Как и ожидалось, у скрытого эталона оценки близки к 100. Для фрагментов 7 («Glockenspiel») и 10 («Человек в длинном черном пальто»), оценки эталона ниже, чем *MP*З при скорости 128 кбит/с, что указывает на «прозрачное» кодирование.

Контрольное прослушивание 2. Оно включает в себя 10 фрагментов ЗС, которые были отобраны для контрольного тестирования кодека MPEG-4 *HE AAC* [2.95]. В тест-программу были включены следующие версии каждого фрагмента:

-оригинал как скрытый эталон;

-прошедший через ФНЧ с полосой пропускания 3.5 кГц;

-прошедший через ФНЧ с полосой пропускания 7 кГц;

-прошедший кодек *aacPlus* (*HE-AAC*) со скоростью пердачи цифровых аудиоданных 24 кбит/с;

прошедший кодек *aacPlus* (*HE-AAC*) со скоростью пердачи цифровых аудиоданных 32 кбит/с;

-прошедший расширенный кодек *aacPlus* (*HE-AAC/PS*) с общей скоростью передачи цифрвых аудиоданных равной 24 кбит/с. При этом использовались 20 полос анализа и не передавались параметры IPD и OPD. Средняя скорость обновления параметров составляла 46 миллисекунд. Для каждого фрейма вычислялось требуемое количество бит для кодирования параметров стерео. Остальные биты были доступны для моно кодера версии *HE-AAC*.

Прослушивание проводилось в два этапа с 8 и 10 опытными слушателями на каждом этапе соответственно. Все отрывки прослушивались через головные телефоны. Результаты каждого этапа, усредненные по множеству отрывков и экспертов, представлены на рис. 2.48. В обоих случаях было

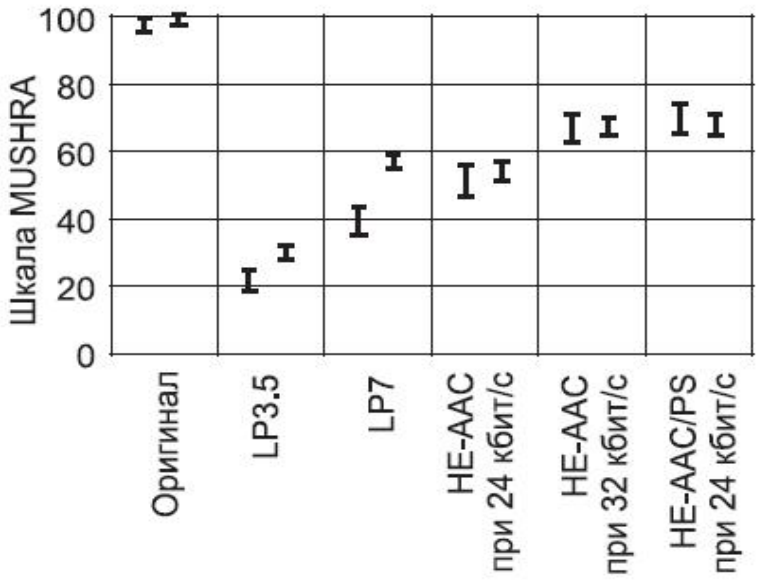


Рис. 2.48. Результаты контрольного прослушивания. Средние результаты при вероятности попадания экспертопоказаний в доверительный интервал равной 95%.: первая позиция слева – качество оригинала; вторая позиция – качество оригинала при его прохождении через ФНЧ с частотой среза 3500 Гц; третья позиция – то же самое, но при частоте среза ФНЧ равной 7000 Гц; позиция 4 – кодек *aacPlus* (*HE-AAC*) при скорости передачи 24 кбит/с; позиция 5 – кодек *aacPlus* (*HE-AAC*) при скорости передачи 32 кбит/с; позиция 6 – кодек *aacPlus* или *HE-AAC/PS)*, когда дополнительно кодируются и передаются на приемную сторону к декодеру пространственные параметры IID, IPD и OPD. Обновление этих параметров выполнялось через 46 мс

выявлено, что кодек *aacPlus* с дополнительной передачей пространственных параметров (улучшенный *aacPlus*) при скорости передачи 24 кбит/с обладает высоким субъективным качеством в среднем около 70% по шкале *MUSHRA*.

Структура декодера в последнем случае имела вид, представленный на рис. 2.46. Итак, применение метода *Parametric Stereo* (PS) позволяет при совместном использовании трех алгоритмов *AAC+SBR+PS* получить при достаточно хорошем качестве суммарное значение скорости цифрового потока 24 кбит/с. При этом полученное качество кодированного сигнала такое же, как и при использовании кодека *HE-AAC* при общей скорости цифрового потока 32 кбит/с. Выйгрыш по скорости составляет около 25%.

2.11. Кодирование сигналов многоканальной стереофонии в стандарте *MPEG D Surround*

Разработка данного стандарта началась в 2004 году, была завершена осенью 2006 года. Здесь так же, как и в алгоритме *Parametric Stereo*, выделяются пространственные параметры, но не двухканального, а уже многоканального звукового сигнала. При этом кодируется их компактный набор, учитывающий свойства слухового восприятия пространственной информации. С помощью переданных пространственных параметров в декодере восстанавливается исходное множество входных сигналов кодера. Кроме пространственных параметров, которые размещаются в аудиофрейме в части дополнительной информации, кодируется также полученный путем матрицирования из исходного множества сигналов двухканальный или одноканальный сигналы. Для передачи пространственных параметров требуется скорость цифрового потока 3…32 кбит/с.

Ряд звуковых сигналов подан на кодер (рис. 2.49). Это могут быть сиг-

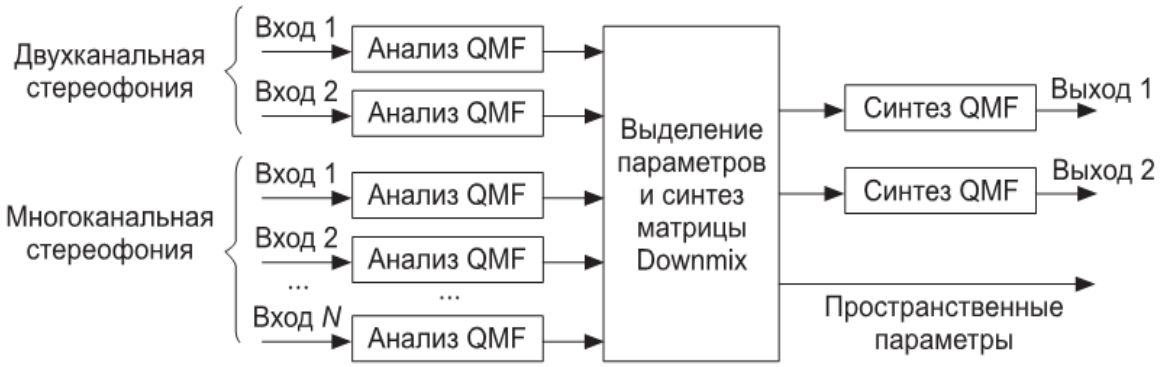


Рис. 2.49. Структурная схема *SAC*-кодера (основное ядро кодера)

налы обычной двухканальной стереофонии (*Input 1* и *Input 2*) или многоканальной стереофонии (*Multi-channel input*) любого формата 5.1; 6.1; 7.1; 10.3;…;22.2 и т.п. Каждый из входных сигналов проходит банк анализирующих квадратурных зеркальных фильтров (*QMF-analysis*) с целью разделения их на полосы частот анализа/кодирования близкие по ширине к

полосам слуховых фильтров (критическим полосам слуха). При разделении спекра входных сигналов на субполосные компоненты не должны возникать слышимые искажения в местах перекрытия (наложения) субполосных сигналов при их объединении в банке синтезирующих *QM*F-фильтров декодера. Кроме того необходимо также, чтобы в декодере была бы возможной интеграция данного метода кодирования с алгоритмами *AAC* и *SBR*, а также и *AAC+SBR* кодера стандарта *MPEG*-4 *HE-AAC* и *AAC+SBR+PS* кодера *HE-AAC/PS*.

Квантование выделенных пространственных параметров, как и в алгоритме *Parametric Stereo*, неравномерное. При малых значениях пространственных параметров шаг квантования мал, при больших значениях – существенно больше. Кроме выделения пространственных параметров (*Parameter estimation*) выполняется также уменьшение размерности исходного многоканального сигнала путем понижающего матрицирования в блоке *Downmix,* обычно до двухканального (обычная стереофония) или одноканального (моно) варианта.

При анализе многоканального сигнала выделяются четыре основных пространственных параметра: различие по уровню (CLD*s*) для каждой пары анализируемых субполосных сигналов; величина корреляции между анализируемыми парами субполосных сигналов; коэффициенты предсказания (CPC*s*); а также сигналы ошибки предсказания или остаточные сигналы. Остаточные сигналы представлят собой ошибку, связанную с матрицированием и в принципе позволяют полностью восстановить исходный многоканальный сигнал. Обычно кодер автоматически генерирует сигнал *downmix*, который оптимизирован для последующего моно- , или для стереовоспроизведения, или для воспроизведения через пассивный декодер звуковой системы *Dolby*.

Структура декодера изображена на рис. 2.50. Она не требует отдельного

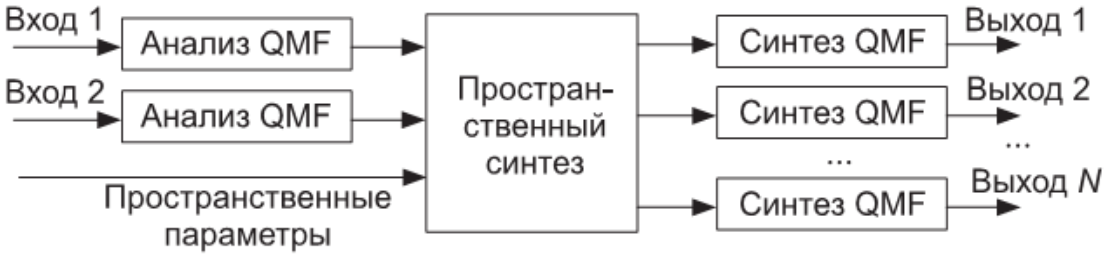


Рис. 2.50. Структурная схема *SAC*-декодера (основное ядро декодера)

пояснения. В нем выполняются обратные преобразования, необходимые для реконструкции входных сигналов кодера.

Число сигналов, подвергаемых кодированию, обычно с помощью алгоритмов *AAC* или *AAC+SBR*, может колебаться от обычного двухканального

формата 2/0 (тогда используемый алгоритм компрессии ничем не отличается от алгоритма *PS*) до формата 10.3 и даже более высокого.

При реализации субполосного анализа/синтеза можно идти несколькими путями:

-построить в лоб нужную систему анализирующих и синтезирующих полосовых фильтров, близких по полосе пропускания к разрешающей способности слуха, однако этот путь встречает при реализации много сложностей;

-выполнить быстрое преобразование Фурье, а затем объединить спектральные компоненты в группы, эквивалентные по частоте разрешающей способности слуха, образовав при объединении критические полосы слуха; данный путь также по ряду причин не эффективен;

-чаще всего для этой цели используют древовидную структуру банков анализирующих/синтезирующих *QMF*-фильтров.

Представленный на рис. 2.51 гибридный банк фильтров содержит на

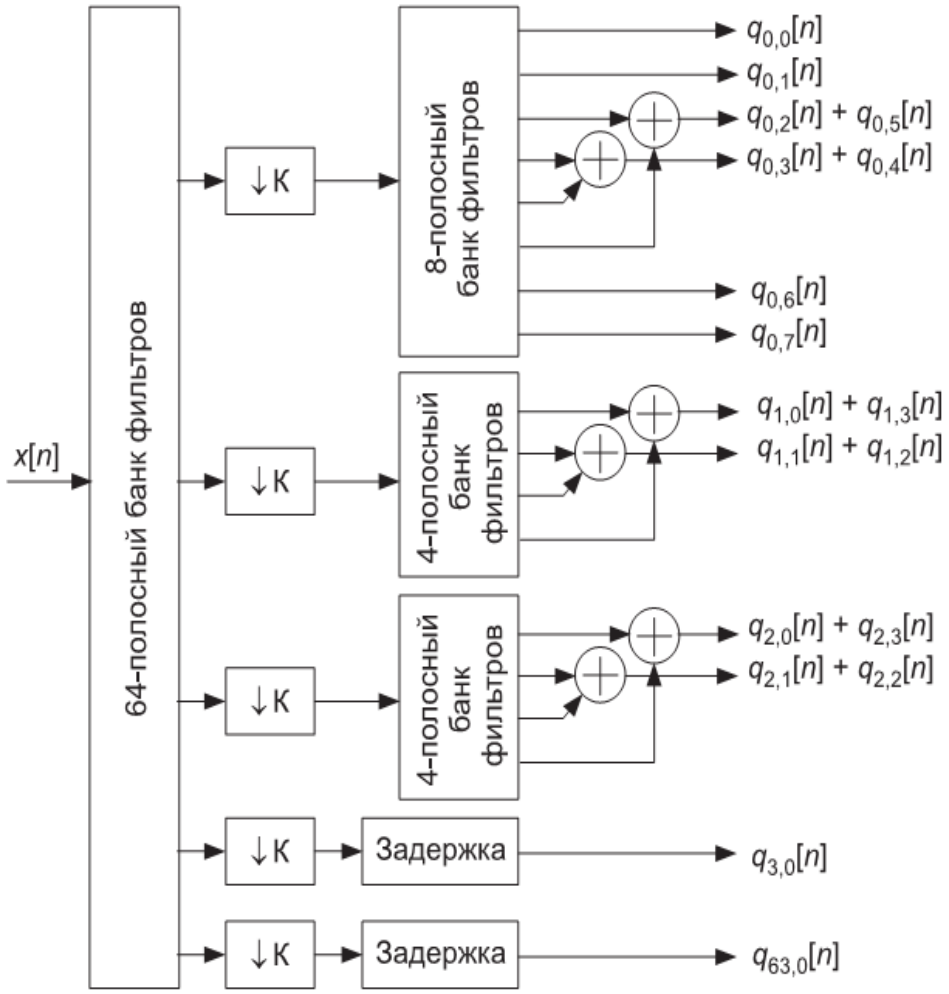


Рис. 2.51. Структурная схема анализирующего гибридного банка фильтров кодера

входе полифазный квадратурный зеркальный фильтр (*PQMF*), разделяющий полосу частот каждого входного сигнала на 64 субполосных компоненты, каждая с полосой частот около 344 Гц. В области низких частот это хуже разрешающей способности слуховой системы человека. Он идентичен банку фильтров, применяемому в алгоритме *AAC* или *AAC+SBR*. Как известно, ширина критических полос слуха на низких частотах составляет около 90…100 Гц. В каждом таком субканале частота дискретизации понижается в 64 раза (К = 64). Далее вторичной системой фильтров наиболее низкие по частоте субполосные компоненты разделяются на еще более узкие субполосы. При этом самая низкая по частоте субполосная компонента разделяется на 8 субполосных (8 *bands analysis filter bank*) составляющих с полосой частот 344:8 = 43 Гц. Две пары наиболее высоких по частоте компонент объединяются здесь, образуя субполосы шириной по 86 Гц. Итак, на выходе данного вторичного фильтра (самый верхний фильтр на рис. 2.51) имеем 4 субполосы анализа каждая шириной по 43 Гц и две субполосы анализа шириной по 86 Гц.

Следующие две субполосных компоненты разделяются вторичным банком фильтров (*4 bands analysis filter bank*) уже на 4 субполосы, каждая 344:4=86 Гц. При этом после их попарного объединения имеем две субполосы анализа шириной по 172 Гц. Таких фильтров два. Полоса частот для остальных субполос остается неизменной, но эти каналы содержат задержку на время необходимое для обработки субполосных компонент в каждом из трех вторичных банков фильтров. Данный банк фильтров имеет разрешающую способность, приближенную к *ERB*-шкале. Подробнее о структуре данного гибридного банка фильтров можно прочитать в [2.43]. В конечном итоге после фильтрации образуются, например, субполосы кодирования, представленные в табл.2.6. Здесь всего имеем 69 субполосных компонент.

Таблица 2.6

Вариант субполос анализа/кодирования

Суб-полоса	Диапазон частот, Гц	Ширина полосы, Гц
0	0...86	86
1	86...172	86
2	172...258	86
3	258...345	86
4	345...517	172
5	517...689	172
6	689...861	172
7	861...1034	172
8–68	1034...22050	345

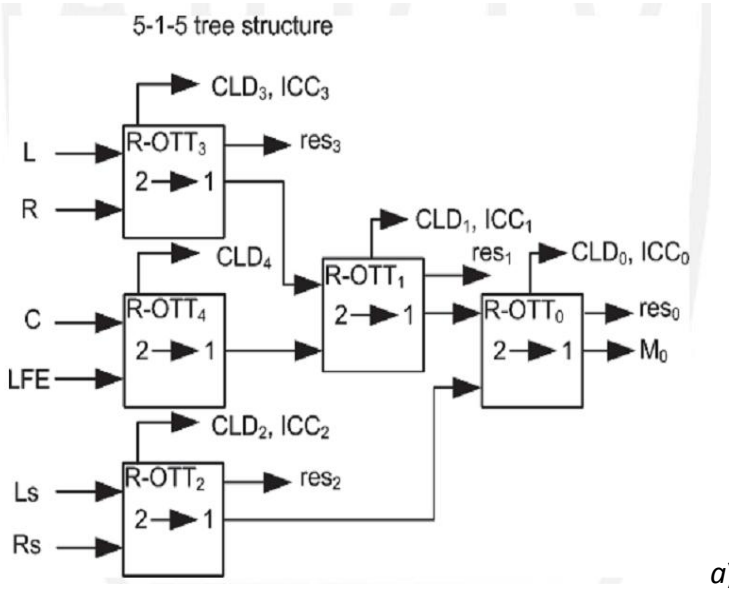
Концептуальные модули OTT и TTT

Кодер *MPEG Surround*. В стандарте *MPEG Surround* используются для извлечения пространственных параметров и для выполненя понижающего матрицирования два типа концептуальных модулей (элементов), получивших название OTT (*one-to-two*) и TTT (*two-to-three*).

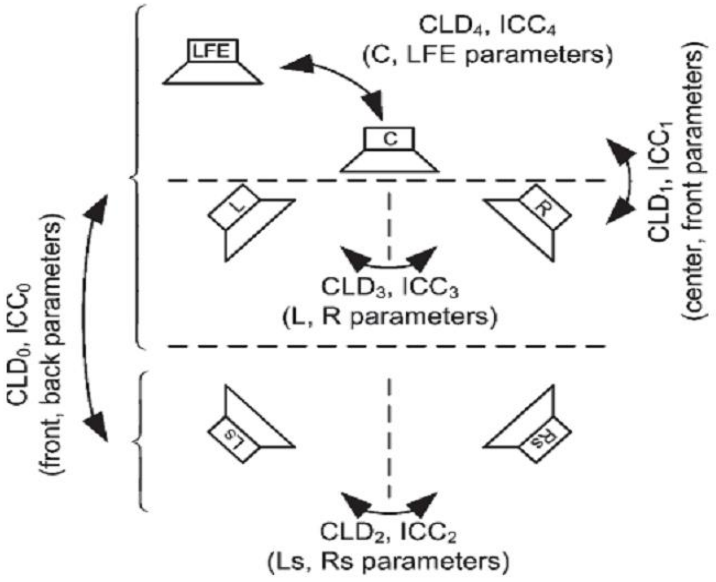
В кодере OTT-модуль извлекает два пространственных параметра: различие по уровню (CLD_S) и значение взамной корреляции IIC_S для каждой пары одинаковых по полосе анализируемых субполосных сигналов. При кодировании этих двух параметрв используется неравномерное квантование с учетом разрешающей способности слуха. В декодере элемент OTT воссоздает два субполосных сигнала из одного, полученного в кодере путем понижающего матрицирования, с помощью переданных от него двух пространственных параметров (CLD_S, IIC_S) и соответствующего повышающего матрицирования.

Элемент TTT кодера создает на выходе из трех входных сигналов путем понижаюего матрицирования двухканальный сигнал, например, L_0 и R_0, пространственные параметры CLD_S и ICC_S и остаточный сигнал res_0^{TTT}. Остаточный сигнал необходим для точного восстановления каждого из трех входных сигналов в декодере. Модуль TTT в декодере из двух сигналов и переданных пространственных параметров CLD_S и ICC_S реконструрирует три исходных сигнала.

Модули OTT и TTT являются базовыми элементами для построения более сложных древовидных структур кодеров и декодеров сигналов многоканальной стереофонии. Примеры таких структур при костроении кодеров и декодеров сигналов многоканальной стереофонии показаны на рис. 2.52 и 2.53.



а)



б)

Рис. 2.52. Древовидная структура кодера формата 5→1/0

Древовидная структура объединяет сигналы формата 5.1 в группы, для каждой такой группы, включающей два сигнала, выполняется понижающее матрицирование, выделяются пространственные параметры и остаточный сигнал. Здесь элементы OTT (верхний рисунок) преобразуют звуковой сигнал формата 5.1 в сигнал формата 1/0 плюс пространственные параметры *Spatial parameters* , выделенные при анализе входных сигналов. В отличие от этого на рис. 2.53 элементы OTT и TTT преобразуют звуковой сигнал формата 5.1 в сигнал формата 2/0 плюс пространственные параметры *Spatial parameters*, выделенные при анализе входных сигналов. Модули OTT используются для выделения пространственных параметров: – различие по уровню CLD и значение коэффициента корреляции ICC (CLD_2, ICC_2, CLD_1, ICC_1, CLD_0), а также остаточные сигналы res_2^{OTT}, res_1^{OTT}. Элемент TTT преобразует три входных сигнала в два, при этом также в процессе субполосного анализа выделяются пространственные параметры CPC/CLD, ICC; в нем путем матрицирования образуются левый L_0 и правый R_0 сигналы стереопары и остаточный сигнал res_0^{TTT} .

При использовании элементов OTT и TTT становится возможным преобразование множества *M* сигналов в *N* сигналов и, наоборот, при условии, что $N < M$. Из модулей OTT и TTT можно строить и более сложные структуры кодеров и декодеров.

Связывая элементы OTT и TTT в различные древовидные структуры можно получить множество конфигураций, например, три из которых представлены на рис. 2.54 для конфигураций 5.1→1/0 (*а*), 5.1→2/0 (*б*) и 7.1 → 5.1 (*в*).

Декодер *MPEG Surround*. Соответствующие древовидные структуры декодеров показаны на рис. 2.55.

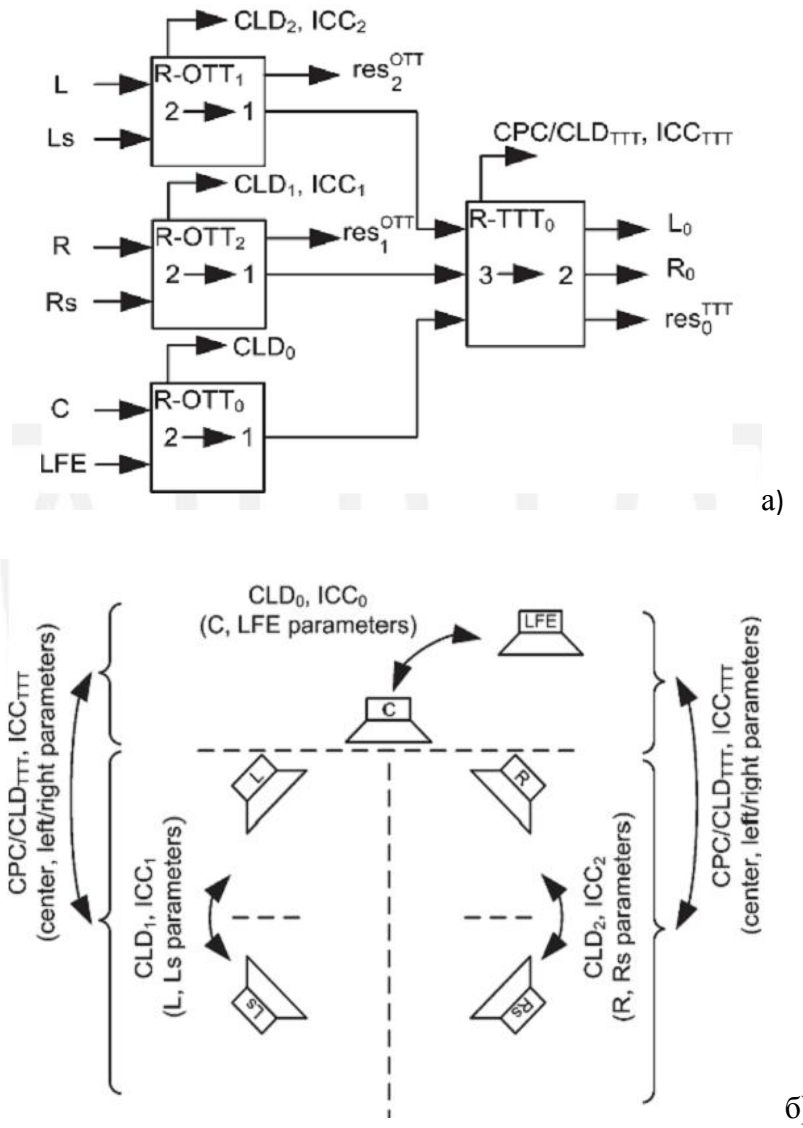


Рис. 2.53. Древовидная структура кодера формата 5→2/0

От передачи декодеру остаточных сигналов res_n^{OTT} res_0^{TTT} часто отказываются. Они представляет собой ошибку моделирования и необходимы при точной реконструкци исходных сигналов в декодере.

Если остаточный сигнал декодеру недоступен, то дополнительно генерируется статистически не связанный сигнал с помощью блока декорреляции (рис. 2.56). Процесс синтеза описывается матрицей смешивания. Передаваемые пространственные параметры: канальная разность уровней (CLD), канальный коэффициент корреляции (ICC), канальный коэффициент предсказания (CPC). Например, элемент OTT берет входной сигнал и

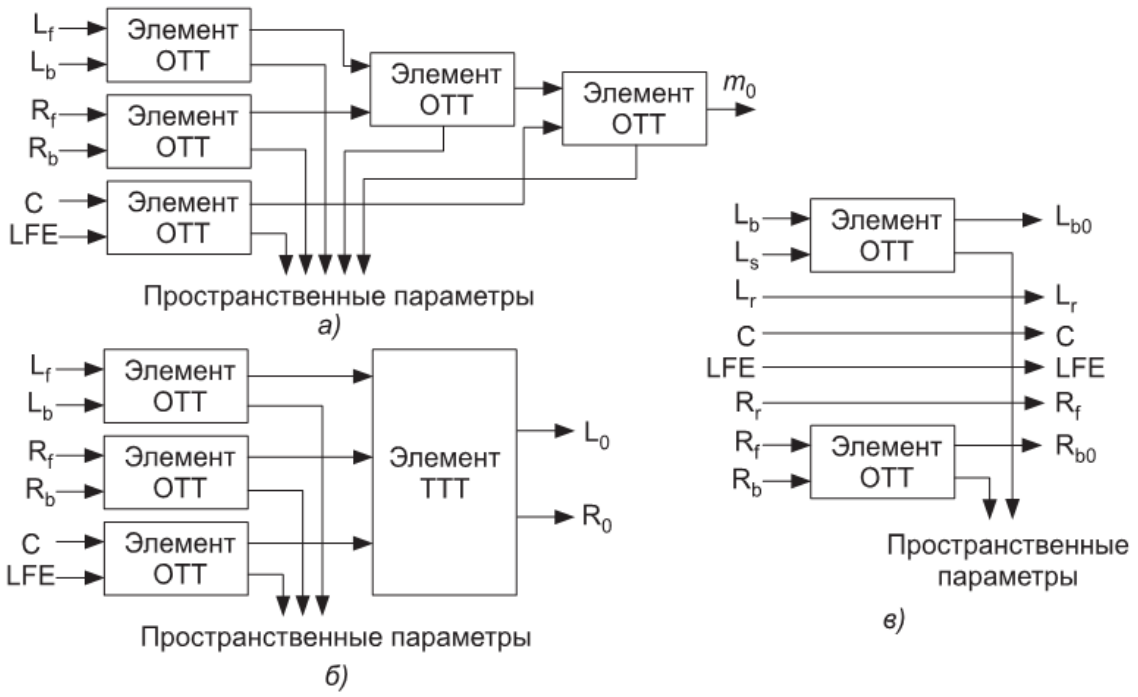


Рис. 2.54. Структура элементов ОТТ и ТТТ для преобразования звукового сигнала формата 5.1 в формат 1/0 (а), формата 5.1 в формат 2/0 (б), формата 7.1 в формат 5.1 (в) плюс компактный набор пространственных параметров (*Spatial parameters*)

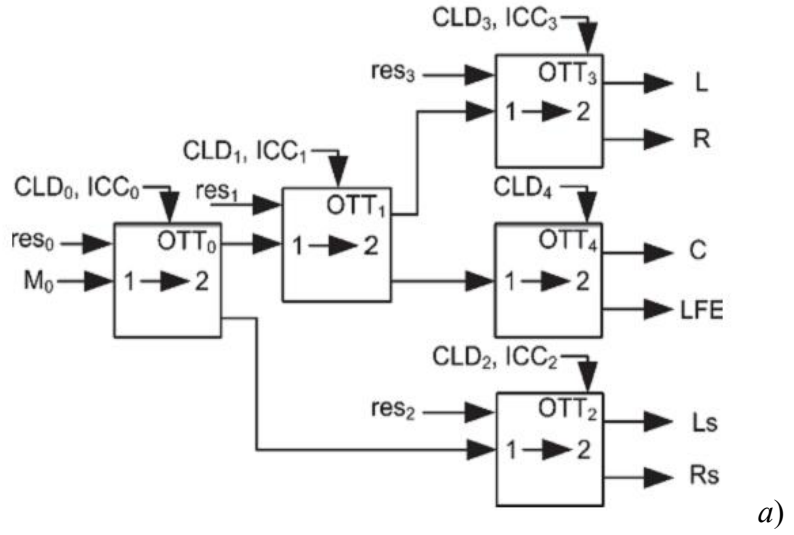


Рис. 2.55,*а*. Структурная схема декодера для реконструкции сигналов формата 5.1 из формата 1/0 с использованием пространственных параметров сигналов формата 5.1

создает его декоррелированную версию с помощью декоррелятора, затем оба сигнала микшируются и с помощью пространственного параметра CLD контролируется распределение энергии в субполосах между ними, а параметр IIC контрлирует декоррелированногосигнала подмешиваемого к каждому из выходных сигналов ОТТ-модуля.

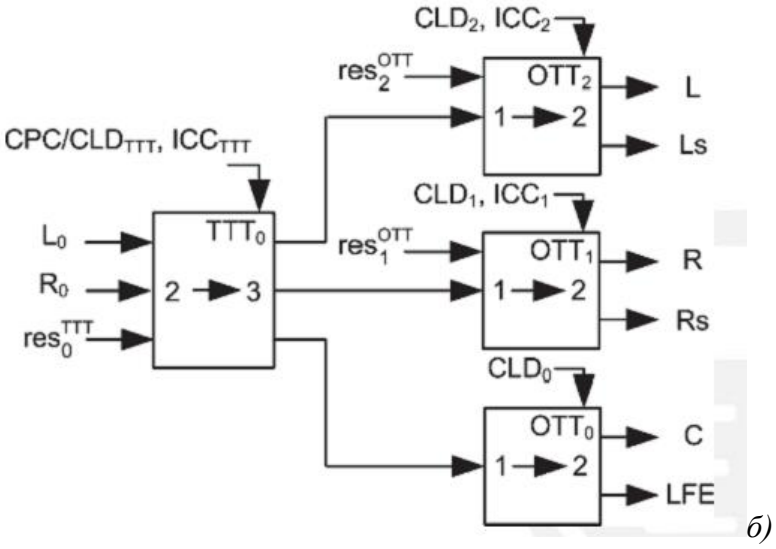


Рис. 2.55,*б*. Структурная схема декодера для реконструкции сигналов формата 5.1 из формата 2/0 с использованием пространственных параметров сигналов формата 5.1

Рис. 2.56. Пример генерации статистически несвязанного сигнала в декодере для получения стереофонического сигнала из монофонического сигнала с помощью декоррелятора

Более сложная структура пространственного синтеза представлена ни рис. 2.57. Она использует две матрицы и несколько декорреляторов для по-

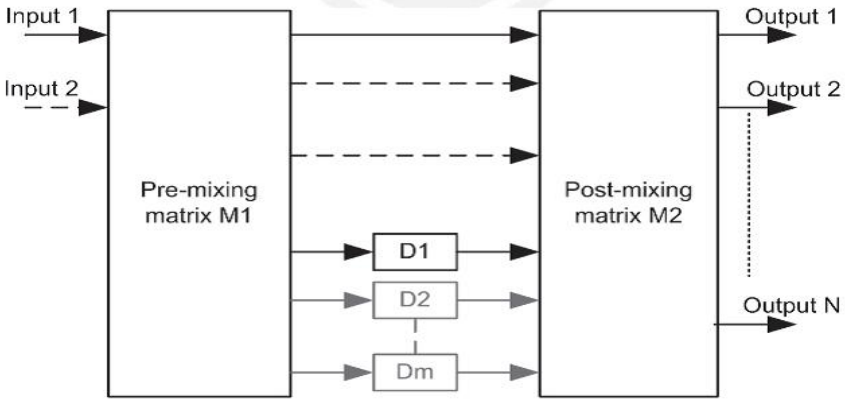
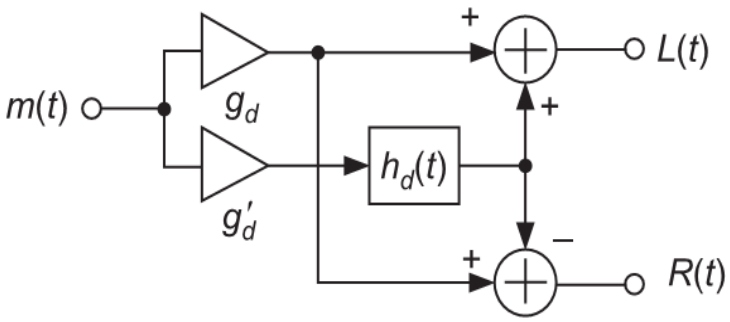


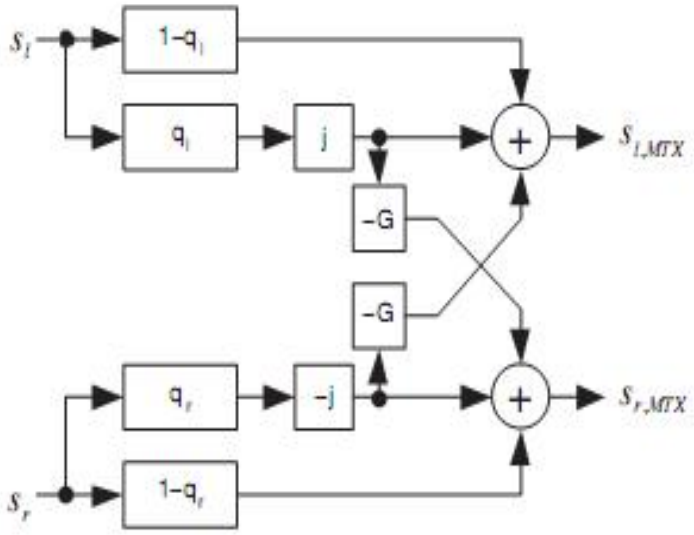
Рис. 2.57. Структурная схема системы синтеза сигналов многоканальной стереофонии с использованием декорреляторов.

лучения многоканального сигнала из двухканального.

Наиболее распространенная базисная структура декоррелятора представлена на рис. 2.58.



а)



б)

Рис. 2.58. Базовая структура декоррелятора: *а* – входной сигнал монофонический; *б* – входной сигнал стереофонический

Можно сказать, что изложенные здесь методы являются развитием алгоритма *PS* применительно к сигналам многоканальных звуковых форматов.

Расширения стандарта *MPEG Surround*

Стандарт *MPEG Surrund* имеет множество детализаций (расширений):

1.Можно менять в очень щироких пределах число субполос анализа/синтеза от 64 до 28, в пределе это число может быть уменьшено даже до одной полосы. Чем меньше число субполос анализа, тем меньшая скорость цифрового потока требуется для передачи пространственных параметров,

но тем хуже будет качество реконструированного декодером многоканального сигнала;

2.Изменение скорости обновления передаваемых пространственных параметров, чем она меньше, тем хуже качество, особенно при передаче быстрых пространственных изменений структуры стерепанорамы;

3.Можно использовать разные разрешения при квантовании самих пространственных параметров. Кроме того предусмотрена также возможность изменения числа передаваемых пространственных параметров. Использование пространственных параметров с низким разрешением осуществляется специальными инструментами, такими как механизм «*адаптивного сглаживания*» параметров (*Adaptive Parameter Smoothing*);

4.Дополнительная передача остаточных сигналов, что позволяет получить качество на уровне дискретных звуковых систем;

5.Данные о межканальной корреляции субполосных сигналов могут содержать только одно значение в целом для всего аудиофрейма.

Применение перечисленных выще опций (расширений) позволяет менять скорость передачи дополнителных параметров в очень щироком диапазоне от 3 до 32 кбит/с.

Декодер *MPEG Surround* может быть реализован в двух версиях: высокого *HQ* и среднего *LP* качества. *LP*-версия требует значительно меньшей вычислительной мощности. При этом обе версии работают с одним и тем же цифровым потоком данных кодера. Версия *HQ* работает с комплексным банком *QMF*-фильтров, а версия *LP* – с вещественным банком фильтров (рис. 2.19). Кроме того версия *LP* испльзует более простые структуры декорреляторов, а операции пред- и постмикширования часто объединяются для повторного микширования некоторых декоррелированных сигналов в процессе синтеза многоканального сигнала.

Матричные технологии в *MPEG Surround*

Помимо понижающего матрицирования исходного многоканального сигнала в форматы 1/0 (моно) или 2/0 (стерео), кодер *MPEG Surround* способен также генерировать с помощью дополнительной матрицы «*matrix-surround*» (MTX) совместимый стереосигнал L_{MTX} и R_{MTX} матричной звуковой системы рис. 2.59). Введение этой процедуры обеспечивает обратную совместимость в декодерах, которые могут обрабатывать только стандартный двухканальный стереопоток *без возможности использования* про-

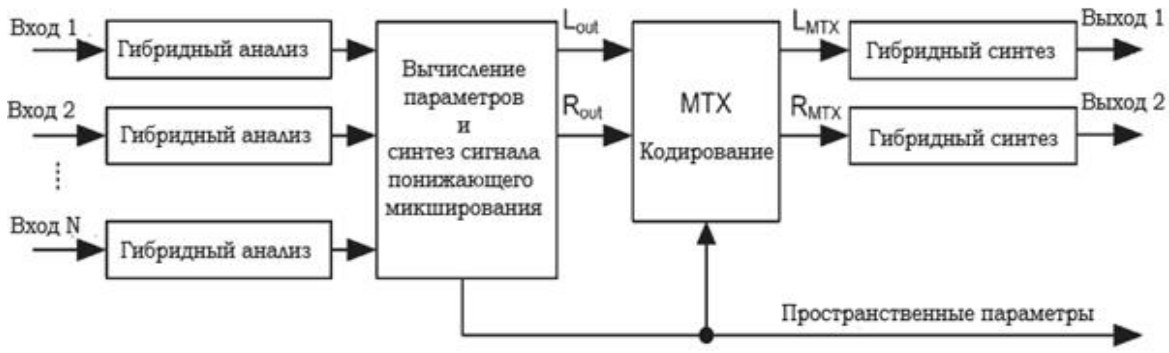


Рис. 2.59. Структурная схема кодера *MPEG Surrund* с матрицей постобработки МТХ

странственных параметров для реконструкции многоканального сигнала формата 5.1. Эта возможность обеспечивается за счет дополнительной процедуры постобработки стереосигнала в матрице МТХ с учетом выделеных в кодере пространственных параметров исходного многоканального сигнала. Обработка сигнала в матрице МТХ выполняется также в субполосах анализа кодера *MPEG Surround*. В дальнейшем пара сигналов $L_{МТХ}$ и $R_{МТХ}$ может быть использована стандартной инверсной матрицей для получения сигнала формата 5.1, как это делается в системах матричной стереофонии с использование преобразования 2→5 (5-2-5). Это так называемый *матричный режим* работы декодера *MPEG Surround*.

В некоторых случаях передача дополнительной пространственной информации невозможна или даже нежелательна. В этом режиме работы не предполагается точная передача пространстсвенной информации и соответственно точная реконструкция исходной стерпанорамы. Это так называемый *расширенный матричный режим (matrix-surround)* работы. В этом случае в кодере генерируется сигнал $L_{МТХ}$ и $R_{МТХ}$ с помощью матрицы МТХ кодера *MPEG Surround* или вообще для этой цели может быть использован стандартный кодер *matrix-surround* матричной звуковой системы. Декодер при этом работает без дополнительных сигналов пространственной информации. Пространственные параметры, необходимые для синеза многоканального сигнала получают прямо в декодере (рис. 2.60). Пространственные параметры – межканальная разность уровней CLD_S и межканальная корреляция ICC_S в субполосах анализа вычисляются для сигналов $L_{МТХ}$ и $R_{МТХ}$, полученных из исходного многоканального сигнала путем понижающего матрицирования. Затем полученные сигналы используются для пространственного синтеза многоканального сигнала. Расширенный матричный режим обспечивает меньшую скорость цифрового потока, но пре-

Рис. 2.60. Декодирование сигналов в режиме «matrix-surround»

восходит по качеству обычные системы матричной стереофонии. При работе в матричных режимах возможно так называемое художественное (творческое) матрицирование, что позволяет слушателям получить оптимальное с их точки зрения качество звучания. Однако, эта технология требует дополнительного изучения.

Стандарт *MPEG Surround* включае также и бинауральные технологии, позволющие реализовать возможности многоканальной стереофонии на портативных устройствах.

Бинауральные технологии в *MPEG Surround*

Одно из самых последних расширений *MPEG Surround* – это возможность создания сигналов бинауральной стереофонии, способных создавать трехмерную звуковую картину при воспроизведении с помощью головных телефонов реконструированной в декодере бинауральной пары сигналов.

Поддерживаются два режима. В режиме *бинаурального декодирования* декодер создает бинауральный двухканальный сигнал, вызывающий ощущения многоканального стерео при воспроизведении через головные телефоны. При втором режиме работы бинауральный сигнал создается в кодере (*режим 3D*), а в декодере этот сигнал может быть преобразован в стандартные многоканальные сигналы для последующего его воспроизведения уже с помощью громкоговорителей. Оба этих способа реализуются на основе новой технологии так называмого бинаурального синтеза с использованием слуховых HRTF-фильтров, реализущих передаточные функции сре-

ды при распространении звуковой волны от источника звука до барабанных перепонок левого и правого ушей слушателя. Напомним, что процесс синтеза бинауральной пары сигналов включает свертку сигнала каждого (виртуального или действительного) источника звука с парой сигналов бинауральных HRTF-фильтров. Использование этой технолгии более подробно изложено в [2.17]. Структурная схема декодера для получения бинауральной пары сигналов при декодировании изображена на рис. 2.61. Демультиплексор раздляет входой цифровой поток на две компоненты:

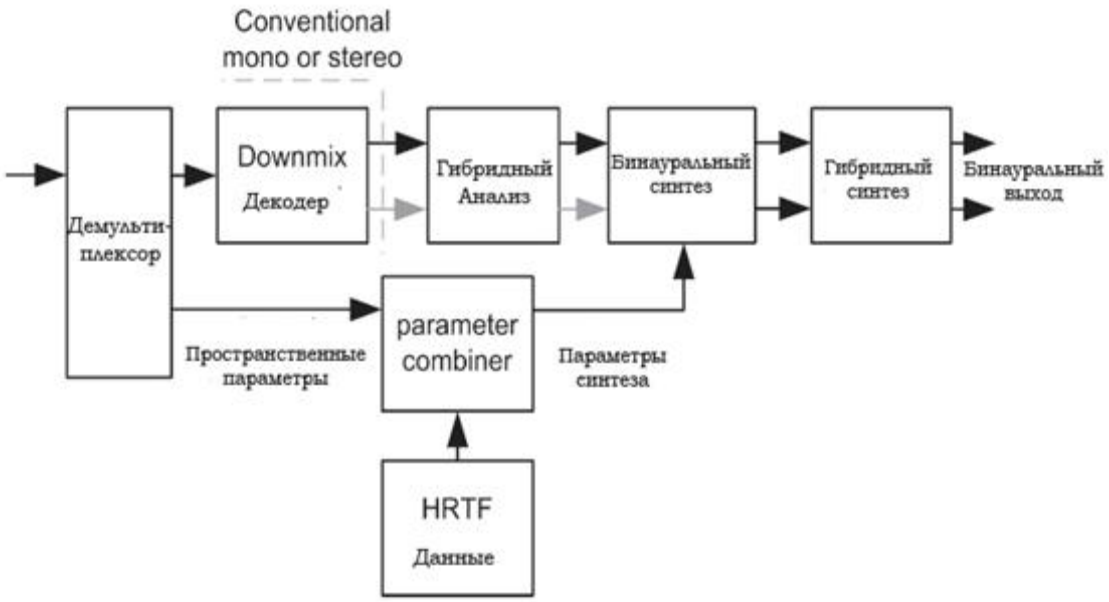


Рис. 2.61. Структурная схема декодера для получения сигналов бинауальной стереофонии (стандарт *MPEG Surround*)

цифровой поток данных микшированного сигнала, полученного в кодере после понижающего матрицирования, и цифровой поток данных, содержащий пространственные параметры исходного многоканального сигнала. Блок бинаурального синтеза также может работать в двух режимах соответственно высокого (*HQ*) и среднего (*LP*) качества. В первом случае используется более точное моделирование характеристик слуховых HRTF-фильтров.

Наиболее простой пример получения бинауральной пары сигналов из любого из имеющихся в нашем распоряжении многоканальных сигналов представлен на рис. 2.62.

При воспроизведении через головные телефоны слушателю кажется, что каждый такой сигнал излучается виртуальным громкоговорителем, вынесенным за пределы головы в точку пространства, определяемую характеристиками соответствуюшей пары слуховых фильтров HRTF(L) и HRTF(R), рис. 2.63). Основной проблемой здесь является синтез слуховых фильтров, требующих достаточно мощных вычислительных ресурсов.

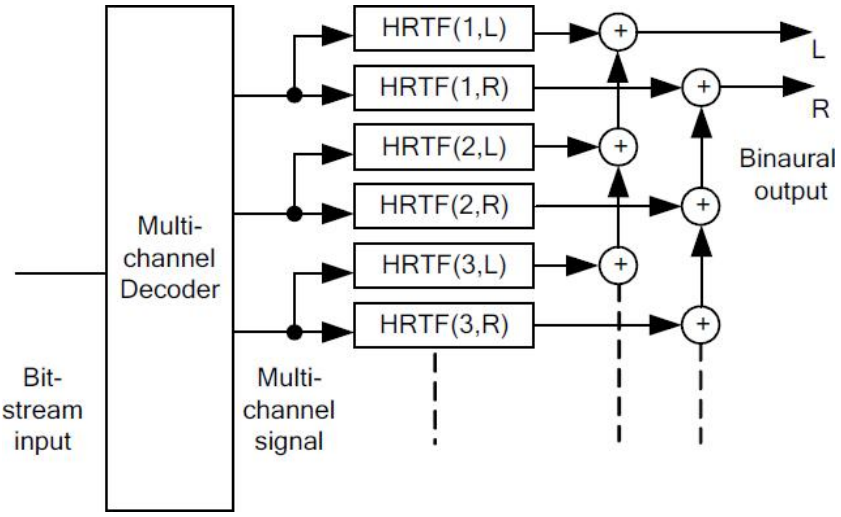


Рис. 2.62. Получение бинауральных сигналов с помощью слуховых HRTF-фильтров

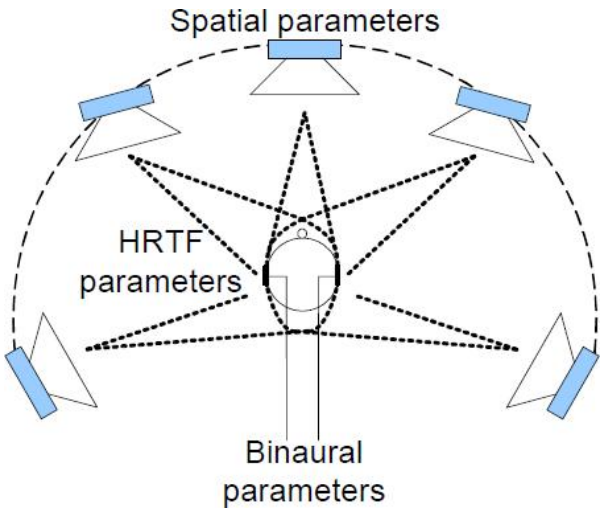


Рис. 2.63. Синтез виртуальных громкоговорителей при воспроизведении бинауральных пар сигналов многоканальной стереофонии

Структурная схема кодера для получения сигнала 3D представлена на рис. 2.64. Схема декодера для его воспроизведения через громоговорители изображена на рис. 2.65. Обе схемы не требуют дополнительных пояснений.

Тестирование бинаурального кодека стандара *MPEG* было выполнено в двух направлениях: оценка возможностей системы по передаче направлений на кажущиеся источники звука и оценка качества кодека в целом в шкале *MUSHRA*.

Локализация КИЗ оценивалась двумя обученными экспертами. Полученные ими результаты представлены на рис. 2.66. На левой группе рисунков углы на источник звука отсчитываются от медианной плоскости головы слушателя в пределах от 0 до ±90 градусов. На правой группе значение угла на КИЗ отсчитывается от фронтального направления по

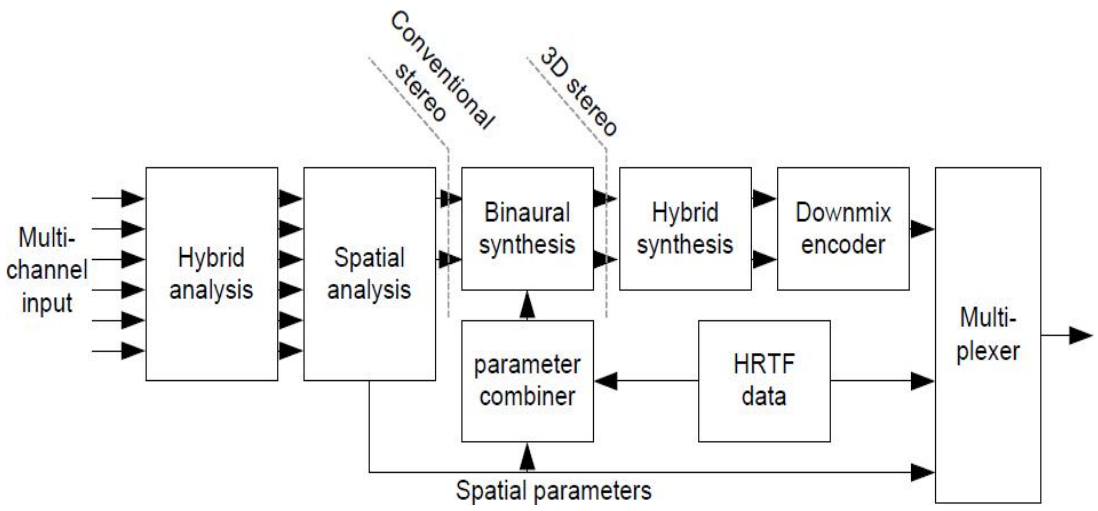


Рис. 2.64. Структурная схема кодера при его работе в режиме 3D

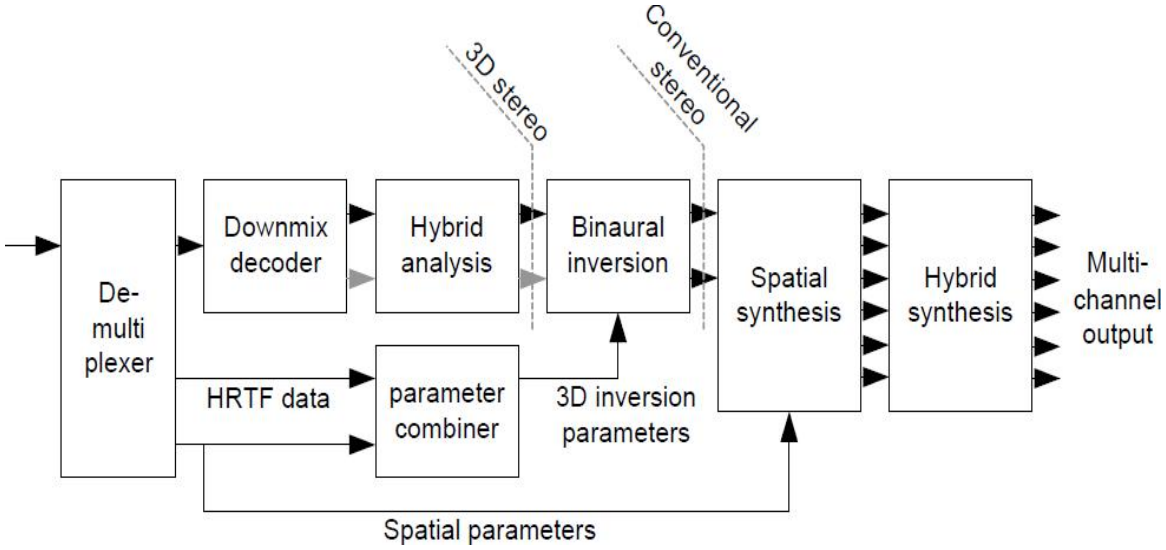


Рис. 2.65. Структурная схема декодера для получения сигнала 3D

часовой стрелке от 0 до 360 градусов. Видно, что при работе в этом режиме сохраняется возможность локализации источников звука в пределах всей азимутальной плоскости. Вертикальные линии показывают разброс показаний экспертов при многокраном повторенииэкспермента.

Результаты оценки качества кодека при бинауральном режиме его работы представлены на рис. 2.67 и 2.68. На этих рисунках по верикальной оси отложены усредненные для всей группы экспертов средние значения оценки качества в шкале *MUSHRA*, в %; по горизонтальной оси – названия отрывков тестовых испытательных сигналов. Верхние данные (Ref) – оценка качества исходного тестового сигнала, самые нижние данные – оценка качества исходного тестового сигнала при ограничении его полосы частот сверху значением 3,5 кГц. В средней части на показаны результаты тестрования кодеков при их работе в двух режимах: бинауральное декоди-

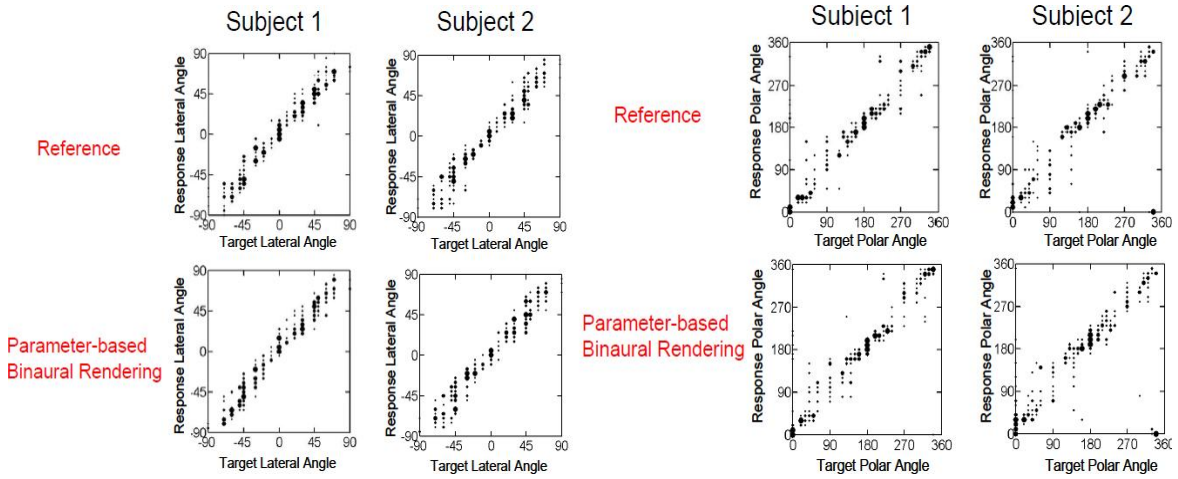


Рис. 2.66. Локализация кажущегося источника звука (бинауральный режим работы кодека стандарта *MPEG Surround*)

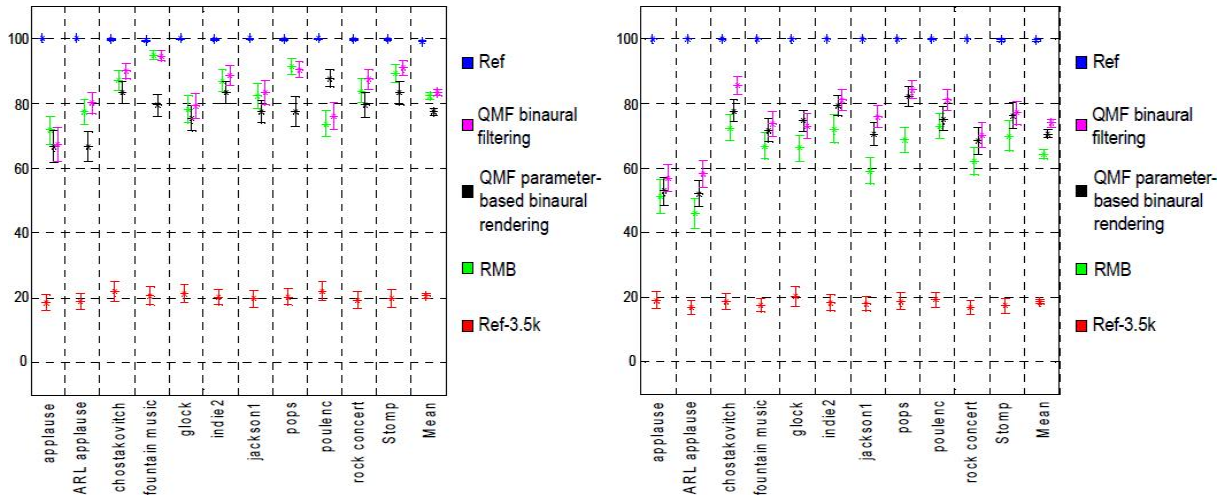


Рис. 2.67. Результаты тестрования бинаурального декодера стандарта *MPEG Surround* в конфигурации ТС-1 (слева) и ТС-3 (справа). Шкала *MUSHRA*

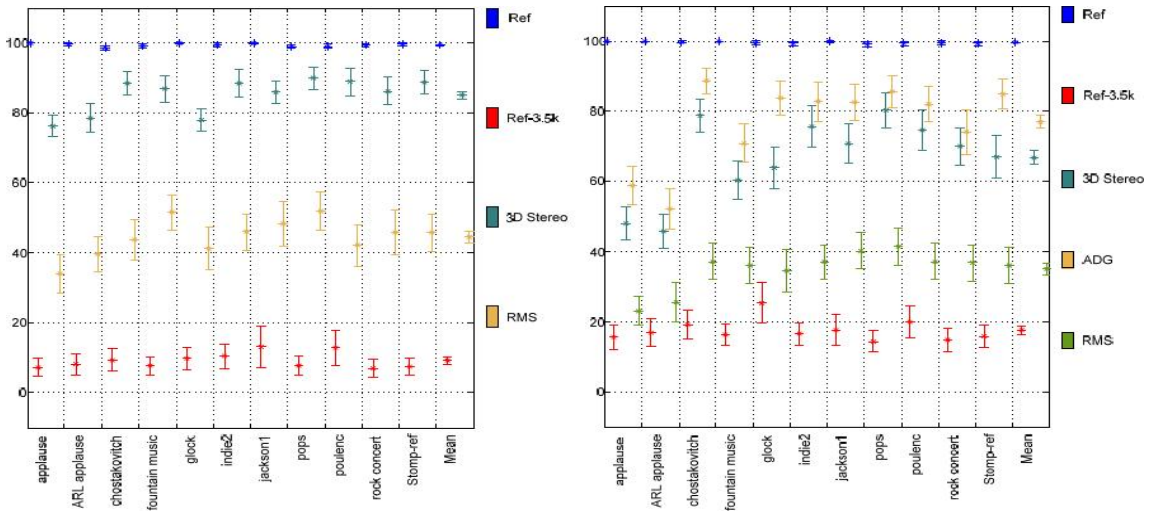


Рис. 2.68. Результаты тестрования бинаурального кодера 3D-стерео стандарта *MPEG Surround* в конфигурации ТС-1 (слева) и ТС-3 (справа). Шкала *MUSHRA*

рование (рис. 2.67) и бинауральное кодирование в формате 3D (рис. 2.68). В конфигурации ТС-1 алгоритм компрессии сигналов AAC, скорость цифрового потока при кодировании сигнала 160 кбит/с. В конфигурации ТС-3 алгоритм компрессии бинауральной пары сигналов *HE-AAC*, скорость цифрового потока 48 кбит/с.

Обозначения на графиках соотвествуют табл. 2.7.

Таблица 2.7

Тестируемые режимы работы кодека

Label	Description
Ref	Original 5.1 item downmixed to binaural with common HRTF set
Ref-3.5k	Anchor, 3.5 kHz low-pass filtered reference
RMB	MPEG Surround (RM) decoder 5.1 output downmixed to binaural with common HRTF set
ADG	Ref downmixed to 3D/binaural and core-coded (3D/binaural downmix as Artistic Downmix, see section 3.1)
RMS	MPEG Surround (RM) decoder 5.1 output downmixed to stereo

Хараатеристики тестовых сигналов представлены в табл. 2.8.

Таблица 2.8

Rendering technique	Test case	Core coder Bitrate	Binaural Filter	Number of stimuli	Number of subjects	Number of rejections
Binaural decoder	TC1	AAC 160kbps	BRTFs 1000 taps	8316	84	11
	TC3	HE-AAC 48kbps	HRTFs (KEMAR)	9108	92	8
3D stereo encoder	TC1	AAC 160kbps	BRTFs 1000 taps	4235	77	17
	TC3	HE-AAC 48kbps	HRTFs (KEMAR)	3036	46	7

Сложность применяемой технологии условно представлена на рис. 2.69.

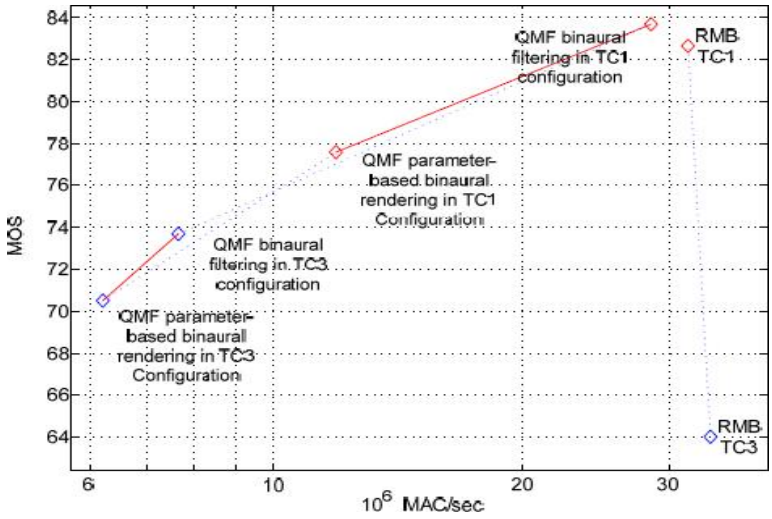


Рис. 2.69. Условное представление сложности технологии

Из изложенного следует, что стандарт *MPEG Surroun*d вобрал в себя все известные техноогии получения сигналов многоканальных дискретных звуковых систем, систем матричной и бинауральной стереофонии, предлагая пользователю право выбора сигналов интересующего его звукового формата из имеющегося их множества. Он пригоден для применения в домашних устройствах, в автомобильном радио, в цифровом радиовещании, обеспечивая качество звучания свойственное многоканальной стереофонии (дискретным, матричным звуковым системам, системам объемного звучания, бинауральным системам).

Самое главное, что стандарт *MPEG Surround* позволяет вести передачу многоканальных сигналов на скорости, близкой к скорости передачи двухканального (или даже в ряде случаев монофонического) звукового сигнала. Также его можно применять ко многим приложениям и сервисам. Результаты тестирования, выполненные независмо разными ведущими фирмами, подтвердили высокое качество, обеспечиваемое этой технологией. Хорошее качество многоканального звука может быть достигнуто даже при очень низких скоростях передачи дополнительной пространственной информации (например, при скорости передачи 3 кбит/с). С другой стороны, использование высоких скоростей передачи пространственых параметров позволяет приблизить качество звука к качеству дискретной многоканальной передачи.

Помимо основной функции кодирования, стандарт *MPEG Surround* обеспечивает множество полезных приложений, которые еще больше увеличивают его привлекательность (например, поддержка художественного микширования, полная совместимость с *matrix-surround* и бинауральное декодирование). Наконец, *MPEG Surround* позволяет использовать многоканальный звук на портативных устройствах из-за малых потерь пространственных данных и возможности воспрозведения «бинаурального» звука.

MPEG Surround идеально подходит для цифрового радиовещания. Учитывая имеющуюся обратную совместимость и малые потери качества при передаче дополнительной пространственной информации, стандарт *MPEG Surround* может быть введен в уже существующие системы цифрового радиовещания без замены обычных устаревших радиоприемников. Они с легкостью будут воспроизводить стереосигнал понижающего микширования, и, если *MPEG Surround* поддерживается, то это позволит приемнику воспроизводить настоящий многоканальный сигнал. Данное тестирование было выполнено для систем цифрового радиовещания (ЦРВ) форматов *DAB, DRM, HD Radio AM, HD Radio FM* и дало положительные результаты.

Качество алгоритмов кодирования пространственных параметров

Как известно, можно, используя только процедуру матрицирования, на передающей стороне, преобразовать любой многоканальный звуковой формат в двухканальный (или одноканальный), затем полученный таким образом сигнал кодировать, а на приемной стороне после его декодирования, используя инверсную матрицу, снова восстановить исходный многоканальный формат. Это один подход (*Matrixed Surround*), используемый в системах матричной стереофонии, например, в системах *Dolby Surround* и *Dolby Pro Logic*. Второй подход состоит в том, что в кодере дополнительно выделяются, квантуются, кодируются и передаются декодеру также и пространственные параметры *SAC*, которые затем используются декодером для синтеза многоканального сигнала. При этом общая картина изменения качества представлена ни рис. 2.70. Из данного рисунка следует, что передача пространственных параметров способна существенно повысить качество

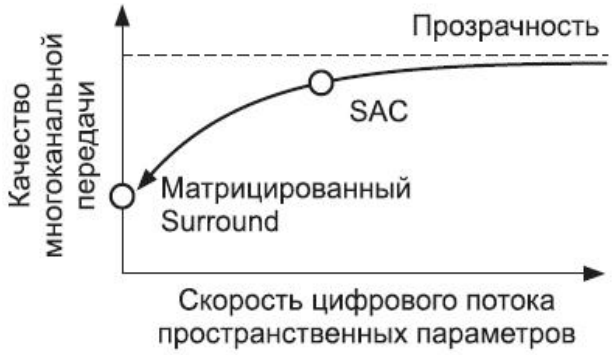


Рис. 2.70. Качество кодирования сигналов многоканальной стереофонии

кодирования сигналов многоканальной стереофонии при одновременном снижении скорости цифрового потока. Этот вывод подтверждают результаты экспертиз, представленные на рис. 2.71. Представленные здесь результаты подтверждают высокую эффективность кодеков при дополнительной передаче декодеру пространственных параметров сигналов многоканальной стереофонии.

2.12. Компрессия цифровых аудиоданных в системах *Dolby Digital*

В системе *ATSC Dolby AC-3* формата 5.1 используется алгоритм компрессии звуковых цифровых данных *A*/52. Он предназначен для кодирования ЗС многоканальной стереофонии, сама же система *ATSC Dolby AC-3* рекомендована национальным комитетом *ATSC* (*Advanced Television System Committee*) США для систем телевидения высокой четкости *HDTV* и

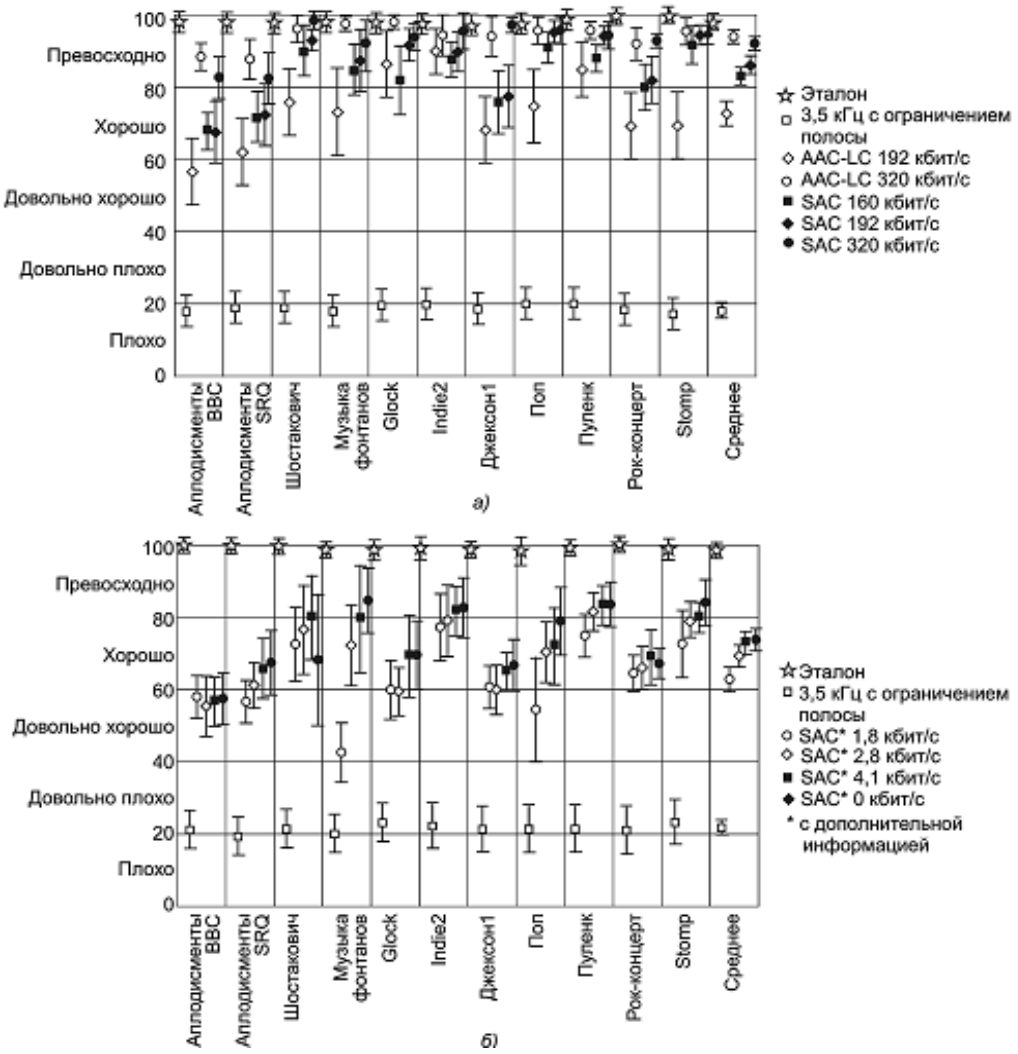


Рис. 2.71. Результаты экспертных оценок высококачественного кодирования (*HQ*, *а*) и кодирования среднего качества (*LR*, *б*) для разных кодеков и различных значений скорости цифрового потока (шкала оценки *MUSHRA*)

других применений: спутниковое вещание, передача звуковых сигналов по оптоволоконным линиям связи, запись на магнитные, оптические и другие носители информации.

Кодер системы *Dolby AC*-3

Кодер системы *Dolby AC-3* предназначен для кодирования высококачественных звуковых сигналов различных форматов от 1/0 (моно) до 5.1. При формате 5.1 по каналам связи в едином цифровом потоке передаются левый *Left,* правый *Reft*, центральный *Center* фронтальные, а также левый *Left Surround* и правый *Reft Surround* тыловые пространственные сигналы и дополнительный сигнал канала сверхнизких частот СНЧ (*Low Frequency*).

В это число включены также форматы 2/0 (обычное стерео), 3/1 и 3/2 (*Dolby-Stereo, Dolby-Surround, Dolby-Pro-Logic*), а также и форматы 3/0, 2/2, 2/1.

Режимы работы кодека и соответствующие им коды представлены в табл. 2.9.

Таблица 2.9

Режим работы и звуковые форматы кодека системы *Dolby AC*-3

Код	Звуковой формат	Кол-во каналов	Звуковые сигналы
000	1+1	2	Ch1, Ch2 (два моно)
001	1/0	1	C
010	2/0	2	L, R
011	3/0	3	L, C, R
100	2/1	3	L, R, S
101	3/1	4	L, C, R, S
110	2/2	4	L, R, LS, RS
111	3/2 или 5.1	5	L, C, R, LS, RS

Упрощенная структурная схема кодера системы *Dolby AC*-3 представлена на рис. 2.72. Цифровой поток на выходе кодера представляет собой последовательность аудиофреймов (*Pack AC-3 Frame*). Содержащаяся в нем информация условно может быть разделена на две части: основную (*Main Information*) и дополнительную (*Side Information*).

Аудиофрейм кодера включает 6 аудиоблоков (рис. 2.73). Каждый аудиоблок содержит информацию о 512 отсчетах для каждого из кодируемых звуковых сигналов (*Audio 1, Audio 2,...,Audio n*). Вследствие 50% временного перекрытия в аудиоблок для каждого из сигналов включаются 256 отсчетов предыдущего блока и 256 новых отсчетов. В 6-ти аудиоблоках аудиофрейма общее число обрабатываемых отсчетов для каждого из входных сигналов будет равно 512·6 = 3072. Заметим, что если число кодируемых ЗС равно 5 (формат 3/2), то общее число отсчетов, информация о которых содержится в одном аудиофрейме, составит (512·5)·6 = 15360, однако с учетом 50% временного перекрытия здесь будет лишь 15360:2 = 7680 новых отсчетов.

После сегментации по времени выборки отсчетов ЗС каждого канала преобразуются в новую совокупность цифровых данных посредством прямого модифицированного дискретного косинусного преобразования (МДКП). Сегментация ЗС по времени с 50% перекрытием выборок и их преобразование из временной в частотную область выполняются в блоке время-частотного преобразования (*Frequency Domain Transform*, рис. 2.72). Перед ортогональным преобразованием выборки отсчетов звуковых сигна-

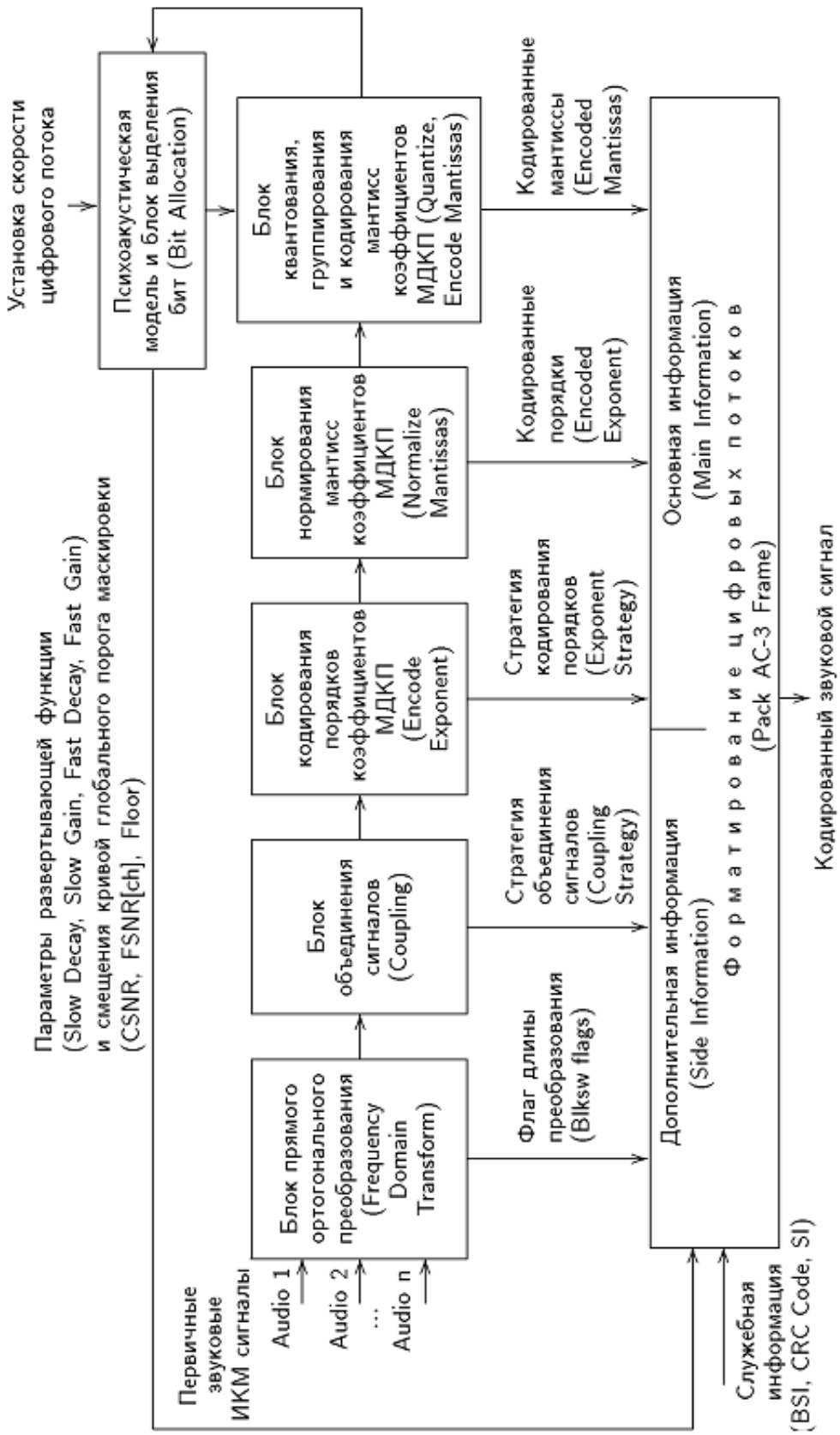


Рис. 2.72. Структурная схема кодера системы *Dolby AC*-3

SI	BSI	Audio Block 0	Audio Block 1	Audio Block 2	Audio Block 3	Audio Block 4	Audio Block 5	AUX Data	CRC

Рис. 2.73. Структура данных аудиофрейма системы *Dolby AC*-3

лов, взвешиваются оконной функцией. Последняя представлена в стандарте *A*/52 таблицей. Форма оконной функции показана на рис. 2.74.

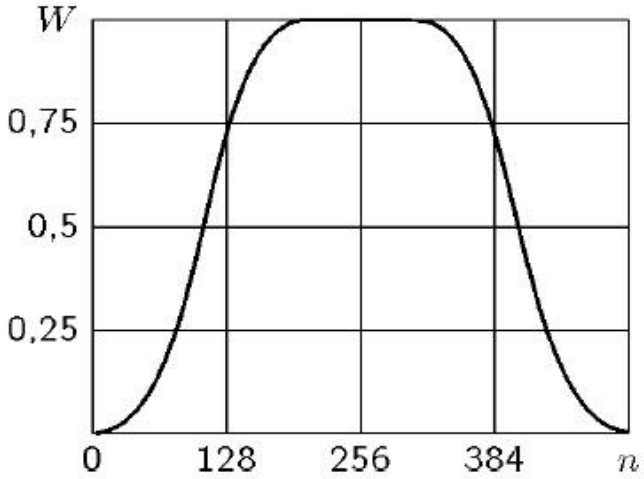


Рис. 2.74. Форма оконной функции

Преобразование выборки ЗС из временной области может быть выполнено посредством одного длинного (512-ти точечного) или двух коротких (256-ти точечных) преобразований. В первом случае будет получено 256, а во втором случае – соответственно 128+128 значений коэффициентов МДКП. При короткой выборке коэффициенты МДКП обоих сегментов, содержащие по 128 значений, объединяются в один общий блок путем их чередования. В этом общем блоке будет также 256 коэффициентов МДКП.

Расчет коэффициентов МДКП проводится по формуле:

$$X_D[k] = \frac{-2}{N}\sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi}{4N}(2n+1)(2k+1)+\frac{\pi}{4}(2k+1)(1+\alpha)\right)$$

при $0 <= k < N/2$,

где N – длина выборки звукового сигнала, $X_D[k]$ - значение к-того коэффициента МДКП; *x[n]* - значение *n*-го отсчета сигнала в выборке, α- параметр равный

$$\alpha = \begin{cases} -1 & \text{для первого (из двух) короткого преобразования,} \\ 0 & \text{для длинного преобразования,} \\ +1 & \text{для второго короткого преобразования.} \end{cases}$$

Заметим, что длинное преобразование наиболее предпочтительно для сигналов медленно изменяющихся по амплитуде с течением времени. Оно имеет лучшее разрешение по частоте. Короткое преобразование обеспечивает лучшее разрешение по времени и применяется для сигналов, амплитуда которых быстро меняется во времени, например, в области атаки звука. Флаг *Block Switch Flags (blksw flags,* рис. 2.72) указывает, какое преобразование (длинное или короткое) применено при расчете коэффициентов МДКП. Параметр *Block Switch Flags* включается в выходной поток цифровых данных как дополнительная информация и используется декодером при выполнении обратного ортогонального преобразования.

При малых скоростях передачи цифровых данных в кодере *Dolby AC*-3 предусмотрено использование специальной процедуры объединения канальных сигналов (*Coupling,* рис. 2.72), позволяющей при их кодировании обойтись меньшим количеством бит.

В системе *Dolby AC*-3 каждый коэффициент МДКП представляется в формате с плавающей запятой двумя значениями: экспонентой (или порядком) и мантиссой:

$$X_D[k] = A[k] \cdot 2^{-B[k]},$$

где A[k] и B[k] – соответственно мантисса и порядок k-того коэффициента преобразования. Порядок равен числу нулей перед первой единицей двоичного представления коэффициента МДКП. Он является по сути дела его масштабным коэффициентом (или нормирующим множителем). Например, если значение коэффициента МДКП $X_D[k]$ = 0,158 и его двоичное представление записывается как 0,001010000110, то значение порядка (масштабного коэффициента) равно B[k]=2, а его мантисса равна 0,1010000110 (в двоичной) или A[k] = 0,6308 (в десятичной) системах исчисления. Очевидно, что $X_D[k] = A[k] \cdot 2^{-B[k]} = 0{,}6308 \cdot 2^{-2} = 0{,}158$.

Знак коэффициента МДКП учитывается при кодировании мантиссы. Перед кодированием значения мантисс нормируются (*Normalize Mantissas*). Значения экспонент и мантисс коэффициентов МДКП кодируются отдельно в блоках *Encode Exponent* и *Quantisse, Encode Mantissas*.

В блоке выделения бит (*Bit Allocaton*, рис. 2.72) учитывается эффект маскировки. В основе процедуры выделения бит лежит модель слуха, позволяющая оценить максимально допустимое (пороговое) значение уровня шума, который еще маскируется полезным сигналом в полосе кодирования, и в соответствии с данными этих расчетов выделить при кодировании мантисс коэффициентов МДКП соответствующее число разрядов. Все указан-

ные вычисления выполняются в блоке называемом обычно психоакустической моделью. Описание этой модели приведено в [2.5]. Каждая нормированная мантисса квантуется с числом ступеней квантования, соответствующим числу бит, определенному в модуле *Bit Allocaton*, рис. 2.72.

Кодирование коэффициентов МДКП в системе *Dolby AC*-3

Итак, в системе *Dolby AC*-3 коэффициенты МДКП представлены в формате с плавающей запятой и имеют мантиссу и порядок, значения которых кодируются с использованием разных процедур.

Кодирование порядков. Порядок коэффициента МДКП в кодере *Dolby AC*-3 представляет собой число, изменяющееся в пределах от 0 до 24. Значение порядка B[k] каждого коэффициента МДКП преобразуется в значение PSD[k] для новой шкалы, содержащей 3072 градации. Поэтому кодовое слово порядка должно иметь, по крайней мере, m=5 разрядов. Максимальное значение порядка ограничено числом 24.

Известно, что если спектр выборки ЗС анализируется с помощью банка фильтров, каждый из которых имеет достаточно узкую полосу частот, то разница в уровнях энергии сигнала между соседними фильтрами редко превышает значение 12 дБ. Это обстоятельство учтено при кодировании порядков. При кодировании значений порядков в кодере системы *Dolby AC*-3 применен метод дифференциальной ИКМ, когда кодируется не само значение порядка, а разность между значениями порядков соседних коэффициентов МДКП. Первое значение порядка для сигнала каждого канала в самой первой наиболее низкой по частоте полосе анализа – это всегда 4-х битовое кодовое слово, что соответствует диапазону изменения чисел от 0 до 15. Значение порядка в следующей вверх по частоте полосе анализа определяется как разница между текущим и предыдущим значениями порядков соответствующих коэффициентов МДКП. В кодере *Dolby AC*-3 разрешающая способность дифференциальной ИКМ (дискретность изменения величин порядков) при кодировании ограничена значениями: -2,-1,0,+1,+2. Максимальное изменение значений порядков соседних коэффициентов МДКП составляет ±2, что соответствует ±12 дБ.

Дифференциальные значения порядков коэффициентов МДКП объединяются в группы. Для процедуры группирования используются три возможных стратегии, обозначенные в стандарте как D15, D25 и D45. В стратегии D15 *каждое* дифференциальное значение порядка кодируется отдельно и ему соответствует одно из пяти возможных значений числа M = 0, 1, 2, 3, 4 в потоке цифровых данных. При этом стратегия D15 требует максимального количества бит при кодировании порядков. В стратегии D25 каждая *пара*, а в стратегии D45 уже каждая *четверка* дифференциальных

значений порядков представлены одним значением числа M в потоке цифровых данных.

Процесс кодирования порядков коэффициентов МДКП с использованием стратегии D15 поясняет рис. 2.75. В верхней его части изображены

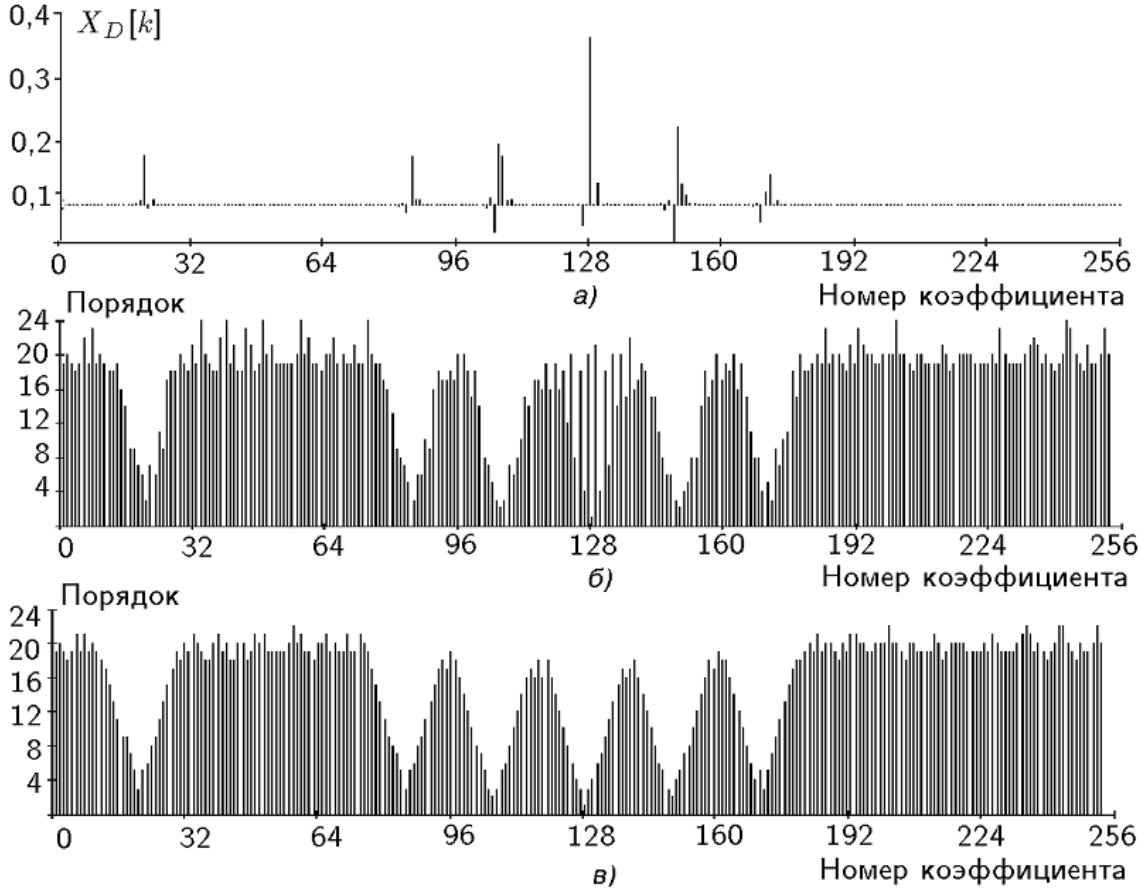


Рис. 2.75. К пояснению стратегии D15 кодирования значений порядков коэффициентов МДКП

коэффициенты МДКП сигнала исходной выборки. В середине и внизу представлены соответствующие им значения порядков соответственно до после их дополнительной обработки.

Дифференциальные значения порядков, полученные непосредственно из исходных значений коэффициентов МДКП (*Original Exponent*), на практике не всегда имеют максимальную величину разности значений соседних коэффициентов, не превышающую диапазон ±2, что требуют соответствующие таблицы стандарта *Dolby AC*-3. Поэтому перед кодированием необходима дополнительная обработка массива значений порядков. С ее помощью уменьшаются некоторые значения порядков, но при этом изменяются и соответствующие им значения мантисс так, что в их двоичном представлении впереди появляются нули. После выполнения этой операции

максимальная дифференциальная величина порядка уже не будет превышать требуемое значение равное ±2.

Выбор стратегии (D15, D25 или D45) кодирования порядков коэффициентов МДКП – это компромисс между хорошим частотным разрешением, разрешением по времени, и количеством бит, требуемых для кодирования экспонент. Стратегии D15 и D25 могут быть использованы для кодирования сигналов, имеющих неравномерный спектр, когда значение экспоненты изменяется довольно быстро от одной полосы анализа к другой. Если же спектр сигнала достаточно гладкий (плоский), тогда используются стратегии кодирования D45.

После выбора стратегии кодирования порядков кодер *Dolby AC*-3 объединяет кодовые слова, соответствующие дифференциальным значениям экспонент, в группы. Для всех режимов работы кодера наборы чисел M для трех соседних (k,k+1,k+2) коэффициентов МДКП M[k], M[k+1], M[k+2] группируются и кодируются как одно 7-битовое слово (*Coded* 7 bit *Grouped Value*) по правилу:

$$M[k,k+1,k+2] = 25 \cdot M[k] + 5 \cdot M[k+1] + M[k+2].$$

Эту процедуру иллюстрирует рис. 2.76 для стратегии кодирования D15. В

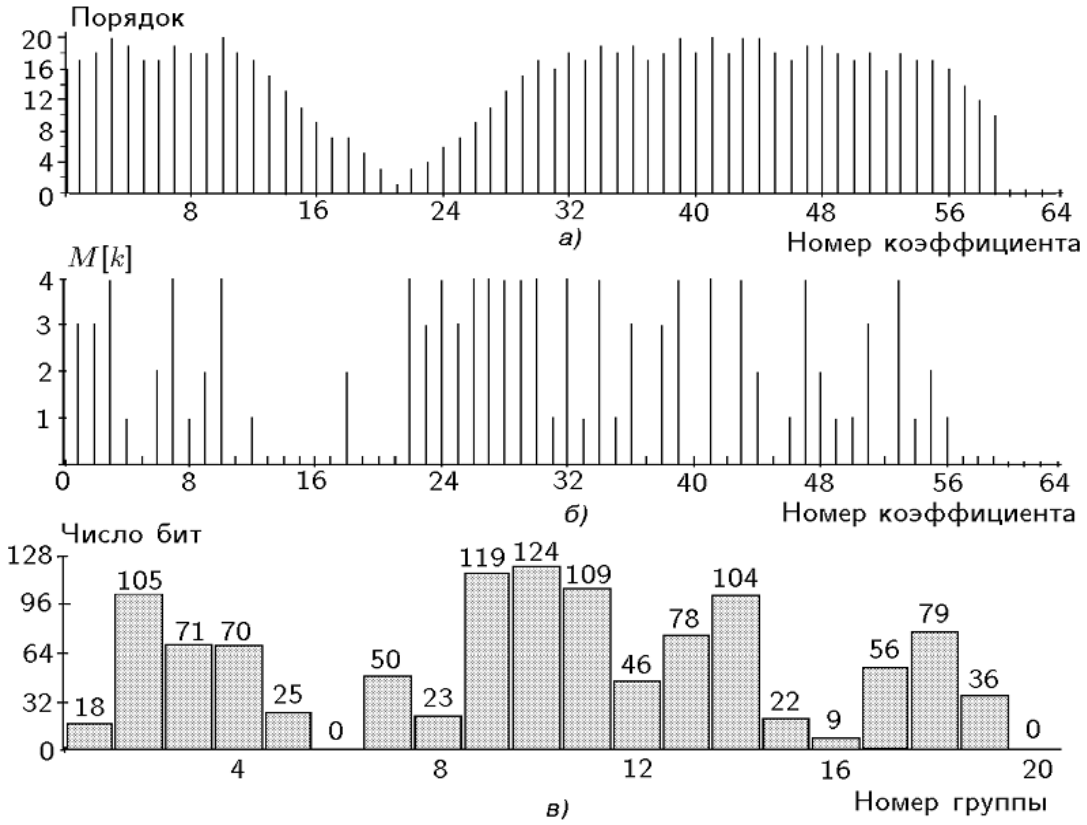


Рис. 2.76. Упаковка дифференциальных значений порядков в поток данных, стратегия кодирования D15

верхней его части показаны значения порядков для коэффициентов МДКП в 63-х полосах психоакустического анализа, в его средней части - значения чисел M[k], а в его нижней части – соответственно число бит, использованное для их кодирования в полосах кодирования после группирования. В общей сложности число полос кодирования равно 20-ти.

Декодер распаковывает и восстанавливает значения порядков, используя инверсную процедуру.

Кодирование мантисс. Диапазон изменения мантисс коэффициентов МДКП лежит в пределах от –1 до +1. Знак коэффициента МДКП учитывается при кодировании мантиссы. При квантовании и кодировании значений мантисс учитываются требования психоакустической модели.

Процесс квантования мантисс коэффициентов МДКП в стандарте *Dolby AC*-3 имеет следующие особенности:

-число возможных ступеней квантования соответствует следующему ряду чисел: 0, 3, 5, 7, 11, 15, 32, 64, 128, 256, 512, 1024, 2048, 4096, 16384, 65536; используется равномерное квантование значений мантисс;

-при числе ступеней квантования равном 3, 5, 7, 11 и 15 используется так называемое симметричное квантование, во всех остальных случаях – асимметричное;

-при числе ступеней квантования равном 3, 5 и 11 кодовые слова мантисс объединяются в группы. При 3-х ступенях квантования три кодовых слова, соответствующие трем значениям мантисс, кодируются одним 5-ти битовым кодовым словом. При 5-ти ступенях квантования 3 кодовых слова мантиссы кодируются одним 7-ми битовым кодовым словом. При 11-ти ступенях квантования два кодовых слова мантиссы кодируются одним 7-ми битовым кодовым словом; в остальных случаях процедуры группирования нет.

При так называемом симметричном квантовании вместо квантованных значений мантисс в цифровой поток включены значения их индексов, заданные соответствующей таблицей. Например, если число ступеней квантования равно 3-м, а значение мантиссы лежит в пределах от –1 до –1/3, то передаваться к декодеру будет значение равное – 2/3 и ему будет соответствовать индекс mc = 0. Если значение мантиссы лежит в интервале от – 1/3 до +1/3, то декодеру передается значение равное нулю и кодируется индекс mc = 1. И, наконец, если значение мантиссы находится в интервале от +1/3 до +1, то декодеру передается значение равное +2/3 и кодируется соответствующий ему табличный индекс mc = 2. Аналогичным образом в форме таблиц задаются интервалы значений мантисс и соответствующие им индексы для числа ступеней квантования равных соответственно 5, 7, 11 и 15. Такой способ квантования позволяет уменьшить число требуемых бит. Для всех других значений числа ступеней квантования (32, 64,…., 65536) кодируются не индексы, а сами значения мантисс коэффициентов МДКП.

Следующим этапом является кодирование и упаковка в цифровой поток значений табличных индексов квантованных мантисс. При симметричном квантовании для уменьшения требуемого для кодирования индексов числа бит используется дополнительно процедура группирования. Например, при числе ступеней квантования равном 7 индекс мантиссы изменяется в пределах от 0 до 6. Для кодирования этого ряда чисел требуется 3 бита. При 11-ти ступенях квантования табличный индекс мантисс лежит в интервале от 0 до 10, а при 15-ти ступенях квантования он находится уже в интервале от 0 до 14. При этом требуемое для кодирования каждого из них число бит соответственно равно 4 или 5. Группирование табличных индексов позволяет уменьшить требуемое для их кодирования число бит при 3-х, 5 и 11 ступенях квантования. При 3-х и 5-ти ступенях квантования три табличных индекса мантисс, а при 11-ти ступенях квантования два табличных индекса мантисс кодируются одним кодовым словом по следующим правилам:

$$\text{Group\_code}[3]= 9\cdot mc[a]+3\cdot mc[b]+mc[c],$$
$$\text{Group\_code}[5]= 25\cdot mc[a]+5\cdot mc[b]+mc[c],$$
$$\text{Group\_code}[11]= 11\cdot mc[a]+ mc[b],$$

где Group_code[3], Group_code[5] и Group_code[11] – кодовые слова групп табличных индексов мантисс соответственно при 3-х, 5-ти и 11-ти ступенях квантования; mc[a], mc[b] и mc[c] – значения табличных индексов мантисс коэффициентов МДКП с номерами a, b и c. Итак, при трех ступенях квантования мантисс (n = 3) кодовое слово группы, состоящей из трех индексов, будет содержать 5 бит, поэтому на кодирование каждой мантиссы в этом случае будет затрачено 5:3 = 1,67 бита. При n = 5 кодовое слово группы мантисс будет представлено уже 7-ми битовым числом и на кодирование каждой мантиссы придется затратить уже 7:3 = 2,33 бита. И, наконец, при n = 11 на кодирование каждой мантиссы потребуется уже 7:2 = 3,5 бита, а при n = 15 на кодирование каждой мантиссы потребуется 4 бита и т. д.

Объединение сигналов стереопары при их кодировании

При работе в этом режиме кодер *Dolby AC*-3 объединяет высокочастотные части исходных сигналов в определенной полосе частот в один общий сигнал и при этом генерирует дополнительно так называемые координаты объединения. Последние будут использованы декодером для восстановления энергетических соотношений высокочастотных частей спектра каждого исходного сигнала, подвергнутого процедуре объединения. После декодирования объединенные части в каждом из восстановленных сигна-

лов будут иметь одинаковый спектральный состав и отличаться только уровнем.

Кодер формирует общий сигнал путем простого сложения коэффициентов МДКП объединяемых сигналов. При этом коэффициенты МДКП с 37-го по 252-ой группируются в 18 субполос (так называемых полос объединения) по 12 коэффициентов в каждой такой субполосе. Нижняя и верхняя частотные границы полос объединения задаются пользователем. Координаты объединения рассчитываются для каждого объединяемого субполосного сигнала. Они представляют собой отношения максимальных значений коэффициентов МДКП каждого объединяемого сигнала и суммарного сигнала в субполосе объединения. Далее координаты объединения преобразовываются в формат чисел с плавающей запятой и включаются в выходной поток данных как дополнительная информация.

Суммарный (объединенный) сигнал кодируется так же, как и сигналы независимых каналов.

Структура аудиоданных в системе *Dolby AC*-3

Структура данных аудиофрейма системы *Dolby AC-3* показана на рис. 2.73.

Поле данных заголовка *(Header)* аудиофрейма содержит информацию о синхронизации *SI* (*Syncronization Information*) и информацию о конфигурации потока данных *BSI* (*Bit Stream Information*).

Поле данных *SI* включает синхрослово (0B77h или 0000 1011 0111 0111), биты помехоустойчивого кодирования (CRC-код), значения частоты дискретизации и размера аудиофрейма. Аудиофрейм системы *Dolby AC*-3 включает два 16-ти битовых слова CRC-кода, первое из них следует в начале каждого фрейма после слова синхронизации, а второе – в его конце (CRC, рис. 2.73). Поле данных *BSI* содержит информацию о конфигурации потока цифровых данных, например, такую как тип сервиса, режим работы кодера (т.е. число кодируемых сигналов или тип звукового формата), абсолютный акустический уровень сигнала каждого канала, информацию о языке, о времени и другое.

Аудиофрейм системы *Dolby AC*-3, как уже было сказано ранее, содержит 6 аудиоблоков. Структура данных аудиоблока показана на

Block Switch Flags	Dither Flags	Dynamic Range Control	Coupling Strategy	Coupling Coordi-nates	Exponent Strategy	Expo-nent	Bit Allocation Parameters	Mantissas

Рис. 2.77. Структура данных аудиоблока системы *Dolby AC*-3

рис. 2.77. Он включает следующие поля битов:

Block Switch Flags – параметр длины ортогонального преобразования;

Dither Flags – признак наличия добавочного шума;

Dynamic Range Control – данные управления динамическим диапазоном передаваемых сигналов;

Coupling Strategy – информация об объединении сигналов (сигналы каких каналов объединены и, начиная с какой частоты);

Coupling Coordinates – координаты объединения для сигнала каждого канала;

Exponent Strategy – выбранная стратегия кодирования порядков;

Exponents– кодовые слова порядков коэффициентов МДКП;

Bit Allocation Parametrs – параметры психоакустической модели;

Mantissas – кодовые слова мантисс коэффициентов МДКП.

В декодере определяется длина кодового слова каждой мантиссы или соответствующего ей табличного индекса, после чего мантиссы распаковываются по специальной процедуре.

Декодер системы *Dolby AC*-3

Декодер системы *Dolby AC*-3 (рис. 2.78) получает форматированный поток цифровых данных (*Input Bit Stream*) и преобразует его в выходные ИКМ-сигналы (*Output PCM*).Первый этап процесса декодирования заключается в распаковке информации аудиофрейма (*Unpack AC-3 Frame*) и разделении ее на основную (*Main Information*) и дополнительную (*Side Informaton*) части.

Декодер *Dolby AC*-3 получает значения порядков коэффициентов МДКП в кодированном и упакованном виде. Чтобы распаковать и декодировать значения порядков, необходимо иметь дополнительную информацию о числе передаваемых экспонент в сигнале каждого канала и о стратегии их кодирования (D15, D25, D45), использовавшейся в кодере. Процесс декодирования порядков осуществляется в блоке декодирования экспонент (*Decode Exponent,* рис. 2.78). После декодирования порядков выполняется процедура распаковки, деквантования и денормирования мантисс коэффициентов МДКП (*Dequantize*, *Denormalize Mantissas*). Для ее выполнения используются параметры психоакустической модели, параметры, определяющие распределение бит в кодере, а также восстановленные значения порядков коэффициентов МДКП. Операция денормирования мантисс производится посредством сдвигов разрядов кодового слова мантиссы вправо. При этом число сдвигов определяется значением соответствующего данному коэффициенту МДКП порядка. Если в кодере была использована процедура объединения сигналов ряда каналов, то, очевидно, что декодер дол-

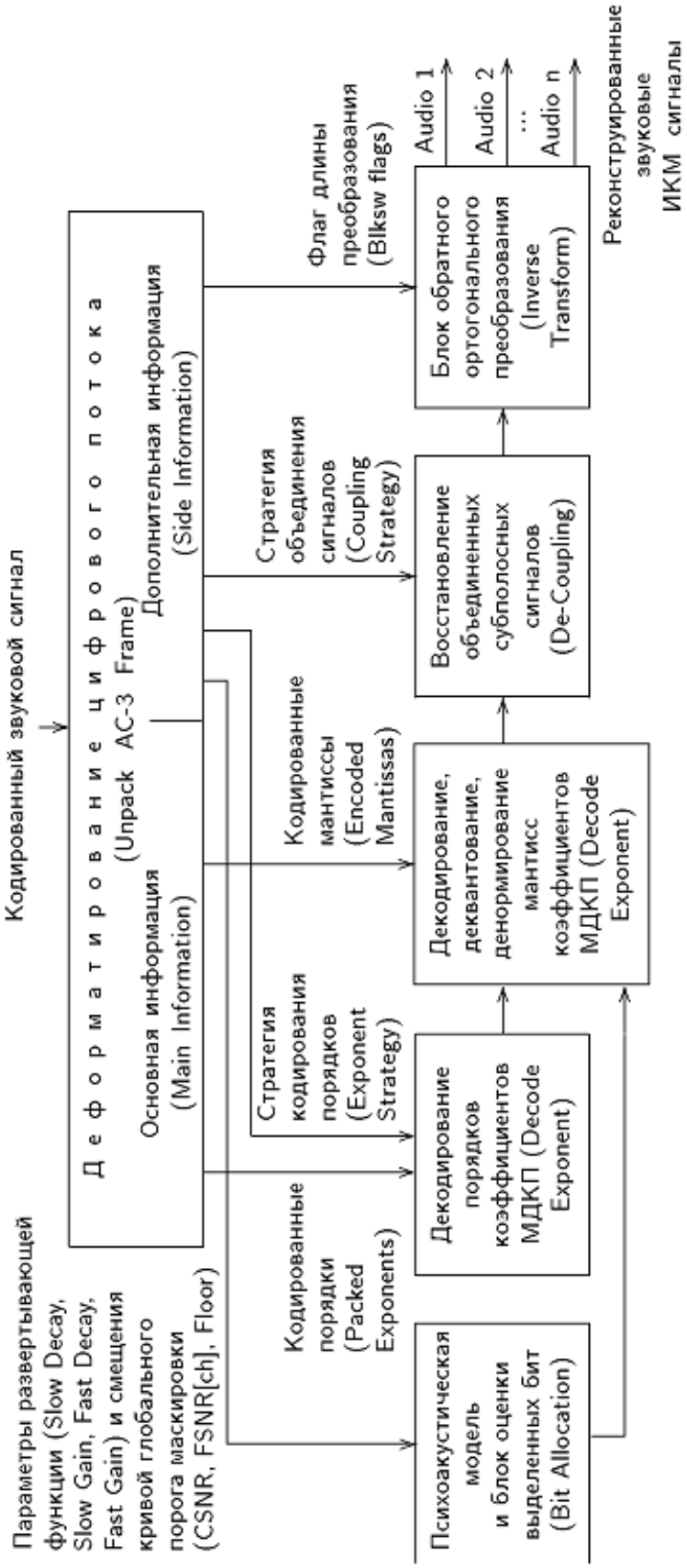


Рис. 2.78. Структурная схема декодера системы *Dolby AC*-3

жен выполнить обратную операцию (*De-Coupling*), используя переданные декодеру в поле данных дополнительной информации значения координат объединения. В блоке обратного ортогонального МДКП (*Inverse Transform*) осуществляется обратное преобразование реконструированного в декодере сигнала во временную область.

2.13. Компрессия цифровых аудиоданных в системе *DTS*

В системе пространственного звучания *DTS* для кодирования звуковых сигналов используется кодек *apt-X*100. В нем применен алгоритм субполосной адаптивной дифференциальной импульсно-кодовой модуляции (АДИКМ) или в английском написании *Subband-ADPCM* (*Adaptive Differential Pulse Code Modulation)*. Напомним, что алгоритм *ADPCM* широко используется для сжатия речевых сигналов. В частности, он рекомендован стандартом *G*.726 (принят в 1984 году) для применения в речевых кодеках. Данный алгоритм обеспечивает качество кодированной речи при скорости цифрового потока равной 32 кбит/с практически такое же, как и при ИКМ и скорости потока равной 64 кбит/с, то есть обеспечивает ее уменьшение в 2 раза. Эффективность алгоритма *ADPCM* повышается еще более при разделении сигнала на полосы, что и реализовано в кодере *apt-X*100.

Входной ИКМ-сигнал имеет в кодере системы *DTS* обычно частоту дискретизации $f_Д$ = 44,1 кГц и разрешение 16 бит/отсчет. Сжатие цифровых данных здесь равно 4:1, суммарная скорость цифрового потока на выходе *apt-X*100- кодера для пяти каналов звука (*L,C,R,LS,RS*) составляет 882 кбит/с при верхней частоте сигнала равной 20 кГц. В настоящее время известно несколько модификаций цифровых форматов в системе *DTS*, ориентированных на разные области применения. Но принято считать (по крайней мере так об этом заявляют разработчики системы *DTS*), что при компрессии цифровых данных равной 4:1 алгоритм *ap*t-*X*100 обеспечивает так называемое *прозрачное кодирование*. Это значит, что искажения, вызванные процедурой компрессии цифровых данных, по отзывам квалифицированных слушателей не заметны на слух. Достоинствами алгоритма *ADPCM* являются:

-малая чувствительность к цифровым ошибкам;

-возможность многократного переприема по низкой частоте, что важно при редактировании и монтаже фонограмм в процессе их записи, и передачи дополнительной информации со скоростью около 12 кбит/с;

-простота реализации кодера (это устройство низкой сложности) при его работе в реальном масштабе времени.

Кодер *apt-X*100

Основными базовыми процедурами системы кодирования *apt-X*100 являются: предварительное разделение спектра исходного звукового сигнала на субполосные составляющие, линейное предсказание; адаптивное квантование и кодирование сигнала ошибки в каждой из выделенных субполос независимо друг от друга.

Укрупненная структурная схема двухканального кодека *apt-X*100 представлена на рис. 2.79. Суммарный цифровой поток левого Л и правого

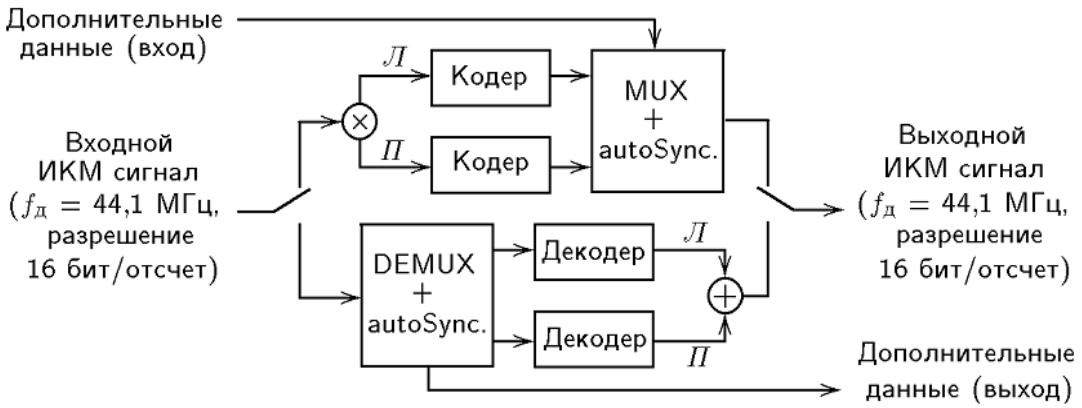


Рис. 2.79. Упрощенная структурная схема двухканального кодека *apt-X*100

П сигналов стереопары разделяется на две части, каждая из которых затем кодируется независимо и после этого мультиплексором (MUX) снова объединяется в единый цифровой поток. В декодере выполняются обратные преобразования: сжатый цифровой поток демультиплексируется (DEMUX), затем каждый из полученных сигналов декодируется, после чего два восстановленных сигнала Л и П при необходимости могут быть снова объединены в единый цифровой поток.

Структурная схема кодера *apt-X*100 показана на рис. 2.80. Входной ИКМ-сигнал, обрабатывается временными блоками, каждый из которых состоит из 4-х последовательных отсчетов ЗС. Эти блоки, кодовые слова отсчетов которых содержат еще по 16 бит, обрабатываются в банке цифровых зеркальных квадратурных фильтров (*QMF*-фильтры), с помощью которого входной ИКМ-сигнал разделяется на четыре одинаковых по ширине полосы частот субполосных составляющих: *LF subband 1*; *Lower MF subband 2*; *Higher MF subband 3* и *HF subband 4*. В каждом таком субполосном канале частота дискретизации понижается в 4 раза. Длина выборки входного сигнала при частоте дискретизации 44,1 кГц составляет 2,7 мс, при $f_Д$ = 48 кГц соответственно 2,5 мс. Полосы частот субполосных сигналов, например, при верхней граничной частоте звукового сигнала равной 20000 Гц составляют соответственно 0…5; 5…10; 10…15; 15…20 кГц.

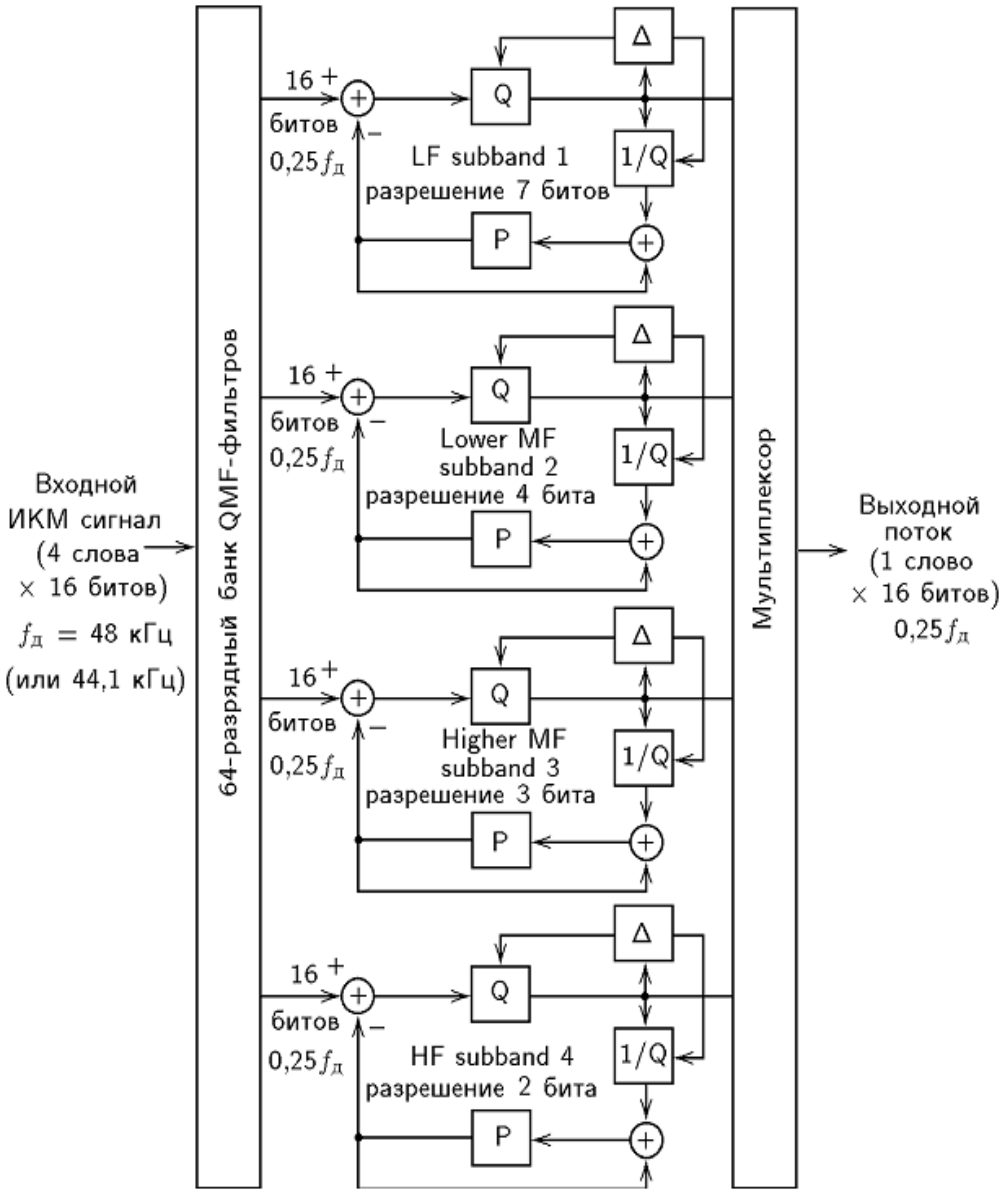


Рис. 2.80.Структурная схема кодера *apt-X*100

На выходах *QMF*-фильтров мы имеем еще 16-битовые кодовые слова отсчетов ЗС.

При разделении звукового сигнала на субполосные составляющие учитываются свойства слуха и спектральные особенности самого сигнала. Напомним, что энергия большинства музыкальных инструментов имеет весьма неоднородное распределение по частоте. Для количественной оценки этого явления часто используют такое понятие как *спектральная неоднородность*, под которой понимается величина показывающая, на сколько спектры реального ЗС и белого шума в субполосе кодирования отличаются друг от друга. Заметим, что струнные музыкальные инструменты (флейта, скрипка и т.п.) создают звучания по своей окраске весьма близкие к тональным сигналам. Их спектры имеют значительную спектральную неод-

нородность, содержат области частот, не играющие существенной роли при слуховом восприятии, то есть они обладают вполне определенной избыточностью. Часто оказывается, что значительная часть энергии сигнала таких музыкальных инструментов содержится в достаточно узких полосах частот, например, вблизи основного тона и некоторых обертонов. В то же время удары тарелок создают сигналы, напоминающие при своем восприятии шум. Они обладают малой спектральной неоднородностью, их энергия распределяется более или менее равномерно на большой диапазон частот. Важно, что для сложных по структуре звука музыкальных инструментов их основной тон расположен в области частот не превышающей 4000 Гц. При этом вне этой области уровень спектральных составляющих достаточно быстро уменьшается. Именно это свойство звуковых сигналов и используется в системе кодирования *apt-X*100. В тех субполосах, где энергия звукового сигнала значительна, их кодирование выполняется с высоким разрешением (длина кодового слова больше). И, наоборот, в тех субполосах, где энергия сигнала минимальна, кодирование выполняется с наименьшим разрешением по уровню. Иначе говоря, при разделении спектра исходного ЗС на полосы и последующем независимом квантовании и кодировании информации в каждой из них учитывается реакция слуха на заметность искажений, вызванных квантованием субполосных сигналов. Это дает определенные преимущества при восприятии, ибо один и тот же уровень шумов квантования неодинаково будет восприниматься слуховой системой человека при субполосном кодировании. Важным достоинством *QMF*-фильтров является также и отсутствие интерференционных искажений в местах стыковки (перекрытия) субполосных сигналов.

Далее отсчеты этих временных блоков после фильтрации обрабатываются в четырех цепях (рис. 2.80), каждая из которых и представляет собой собственно АДИКМ-кодер. Она содержит сумматор (+), квантователь Q, линейный предсказатель P, вычитатель (–), устройство адаптации шага квантования Δ, инверсный квантователь 1/Q. Сигнал, формируемый на выходе предсказателя P в каждый текущий момент времени, учитывает предысторию сигнала: он формируется на основе учета значений 122 предшествующих отсчетов звукового сигнала. Эти 122 отсчета обуславливают величину задержки предсказанного значения по отношении к текущему моменту времени. Текущее и предсказанное значения вычитаются, квантуется и кодируется их разность, что требует существенно меньшего числа бит. Кодовое слово разностного сигнала называется сигналом ошибки, оно еще по-прежнему содержит 16 разрядов. Можно сказать, что сигнал ошибки квантуется повторно с использованием адаптивного квантователя Лапласа. При этом размеры шага квантования изменяются ступенями в зависимости от абсолютной величины сигнала ошибки. Изменение величины шага квантования также базируется на анализе изменения величин предшествующих отсчетов ЗС. В итоге достигается постоянно оптимальное разрешение

квантованного сигнала ошибки, а, следовательно, и преобразование формата сигнала и его сжатие.

Итак, в цепи линейного предсказания текущее значение отсчета ЗС сравнивается с вычисленным предсказанным значением. Очевидно, что предсказанное значение может быть меньше или больше текущего значения отсчета. В каждом случае этот сигнал ошибки вычисляется как разность сравниваемых отсчетов. Если предсказанное значение будет вычислено точно, то уровень сигнала ошибки будет во много раз меньше значения текущего отсчета и его можно повторно квантовать Q с существенно меньшим разрешением, чем исходное 16-ти битовое слово.

Предсказание базируется на значении предшествующего отсчета, которое реконструируется инверсным квантователем (1/Q). При этом, конечно, имеется в виду, что кодер и декодер во всем диапазоне возможных изменений уровня могут генерировать идентичные предсказанные значения при отсутствии какой-либо телекоммуникационной связи между ними. Благодаря этому точные значения редуцированных избыточных частей сигнала в декодере могут быть снова реконструированы. Здесь важно отметить следующее. Эффективность (точность) линейного предсказания растет при наличии в сигнале явной периодичности и благодаря этому свойству может быть существенно повышена, что и реализовано в системе кодирования *apt-X*100. Заметим, что чистые тоны или тонально похожие сигналы воспринимаются с очень высоким разрешением, то есть слух способен их выделять. При наличии в сигнале значительной периодичности генерируемый в цепи линейного предсказания сигнал ошибки очень мал, поэтому кодирование оказывается в этом случае возможным с максимальной точностью (высокая точность предсказания). И, наоборот шумоподобные сигналы не вызывают при слуховом восприятии слишком четких ощущений, их периодичность в сравнении с тональными сигналами незначительна, что является причиной появления большого сигнала ошибки при линейном предсказании. Однако интересно здесь то, что такой сигнал с позиций слухового восприятия может кодироваться с малым разрешением.

Разрешение (число бит, предоставленных для кодирования) квантователя разностного сигнала внутри различных субполос выбирается постоянным по величине и не зависящим от уровня сигнала ошибки. Это линейный квантователь, величина шага которого постоянна во всем диапазоне изменения уровней. В первой из субполос кодирования (рис. 2.80) длина кодового слова составляет 7 бит/отсчет, во второй – 4, в третьей – 3 и в последней – 2 бита/отсчет. Отсчетом здесь служит сигнал ошибки. Итак, в каждой субполосе кодирования независимо от уровня сигнала ошибки последний всегда кодируется с одним и тем же разрешением, то есть кодовые слова имеют одинаковое число разрядов.

При равномерном квантовании возникают определенные трудности. С одной стороны, шаг квантования следует выбирать таким, чтобы диапа-

зон квантователя использовался бы полностью, то есть диапазон квантователя должен быть согласован с размахом сигнала. С другой стороны, шаг квантования следует делать малым для уменьшения искажений (шумов) квантования. Эта еще более усложняется нестационарным характером звукового сигнала, ибо его амплитуда, включая и амплитуду сигнала ошибки, может изменяться в широких пределах. На это влияют факторы уже перечисленные выше. Все это требует адаптации свойств равномерного квантователя в данном случае к уровню сигнала ошибки. Если адаптивное квантование применяется непосредственно к сигналу ошибки, представляющее собой разность исходного и предсказанного значений, то такой метод обработки называется адаптивной дифференциальной импульсно-кодовой модуляцией (АДИКМ). Его идея здесь состоит в том, что число ступеней квантования в субполосе кодирования остается постоянным для любого уровня сигнала ошибки, а его величина шага квантования при этом меняется в соответствии с изменениями уровня последнего так, чтобы для каждого отсчета использовалась бы полностью вся шкала квантователя. Причем (рис. 2.80) в данном случае адаптация шага квантователя выполняется по выходному сигналу, его величина в данном случае зависит лишь от значения предшествующего кодового слова. Предсказанное значение восстанавливается из сигнала ошибки с помощью инверсного квантователя. В итоге выбирается ступенчато такое значение шага квантования, которое минимизирует мощность шумов квантования.

Итак, при АДИКМ величина шага квантователя непрерывно приводится в соответствие с уровнем сигнала, чтобы достигать постоянно минимума шумов квантования. Если энергия сигнала в субполосе остается во времени постоянной, то и величина шага квантования не изменяется. Постоянные колебания уровня, сигнала ошибки уменьшают эффективность квантования. Немаловажную роль при этом играют и эффекты временной маскировки, когда порог слышимости повышается на коротких временных отрезках до и после прихода выброса ЗС.

В результате после процедуры адаптивного квантования четыре 16-ти битовых кодовых слова временного блока (всего 16x4=64 бита) будут уже содержать в сумме только 16 бит (7+4+3+2=16 бит), следовательно, сжатие данных составляет 4:1. Итак, разрешения или число ступеней квантования в каждой субполосе различны и много меньше, чем для входного ИКМ-сигнала. Частоты основных тонов музыкальных инструментов и голосов лежат в нижней субполосе. Здесь разрешение квантователя выше. В области же более высоких частот расположены обертоны, точность кодирования амплитуд которых может быть меньше. В самой верхней субполосе спектр сигнала по форме напоминает шум и для его кодирования требуется наименьшее число бит. Вследствие этого скорость цифрового потока в каждой субполосе различна. В мультиплексоре цифровые потоки субполосных сигналов объединяются в общий цифровой поток, к которому добавляется

также служебная информация, необходимая для правильного его декодирования, и дополнительные данные.

Декодер *apt-X*100

В декодере (рис. 2.81) *apt-X100* выполняются обратные преобразования: редуцированный сигнал преобразуется здесь снова в последовательность 16-ти битовых кодовых слов равномерной ИКМ.

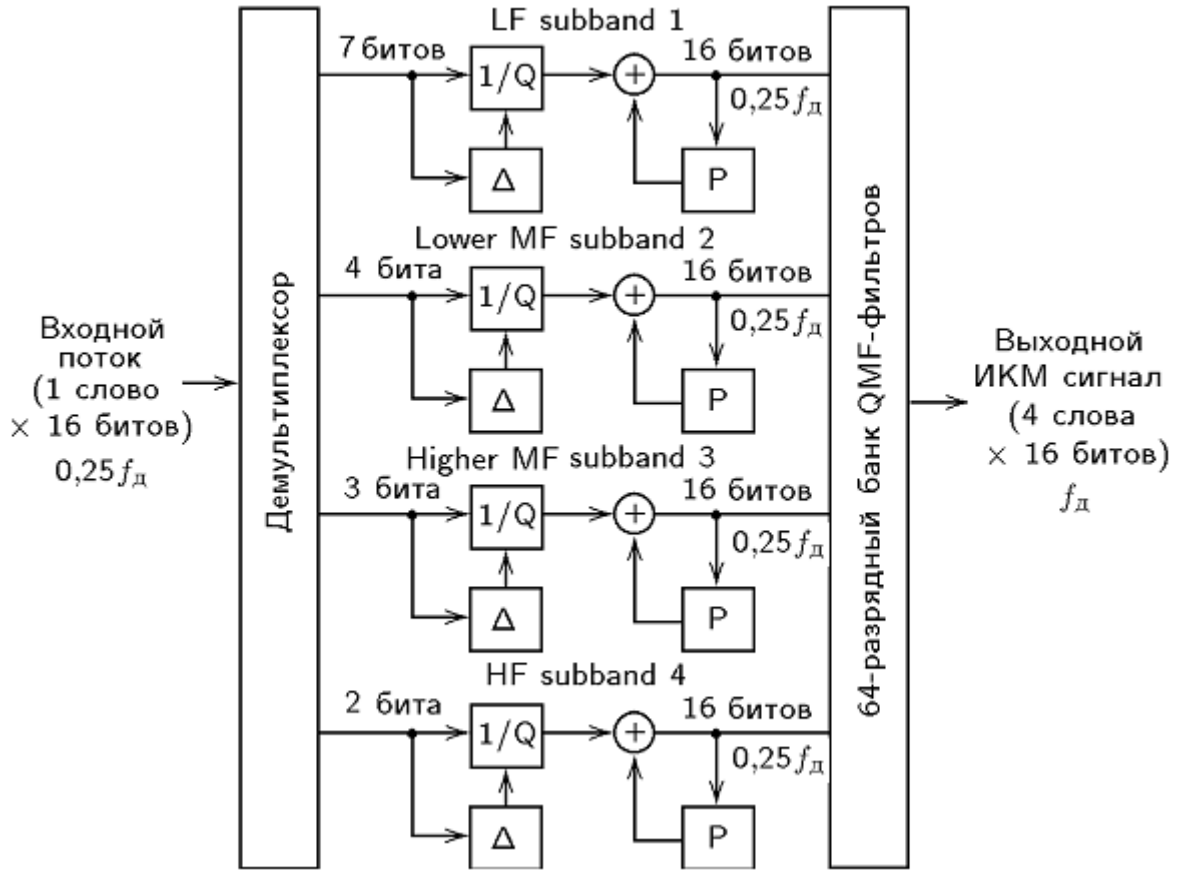


Рис. 2.81. Структурная схема декодера *apt-X*100

Сжатый входной цифровой поток демультиплексируется (*De Multiplexer*). При этом каждый 16-ти битовый временной блок разделяется на четыре компоненты соответственно содержащие 7, 4, 3, 2 бита, каждая из которых направляется в свой (один из четырех) канал обработки, где в результате декодирования и происходит восстановление исходных 16-ти битовых кодовых слов. На выходе инверсных квантователей 1/Q с помощью блока управления величиной масштабного коэффициента Δ восстанавливаются 16-ти битовое кодовые слова каждого из четырех отсчетов сигналов ошибки. Затем каждый из этих сигналов поступает на сумматор и с его вы-

хода на цепь линейного предсказания Р. Предсказанное 16-ти битовое значение текущего отсчета, как и ранее, формируется также на основе 122 предшествующих его значений. В итоге на выходах каждого из сумматоров этих четырех цепей будем иметь восстановленные 16-ти битовые кодовых слова соответствующих субполосных отсчетов. Далее эти восстановленные субполосные сигналы поступают на банк инверсных (синтезирующих) квадратурных зеркальных фильтров (*QMF*-фильтры) где и объединяются в единый цифровой поток, образуя последовательность 16-ти битовых кодовых реконструированного исходного ИКМ-сигнала.

При необходимости сигнал с выхода декодера может быть подан на цифро-аналоговый преобразователь (ЦАП) для получения аналогового сигнала соответствующего канала воспроизведения системы *DTS*.

2.14. Компрессия цифровых аудиоданных в системе *SDDS*

Первоначально алгоритм компрессии *ATRAC-Adaptive TRansform Acoustic Coding* был разработан фирмой *Sony* для системы записи аудиоданных на *MiniDisk* (MD) в 1992 году, когда ею на рынке был представлен первый минидисковый плейер (*MD*-плейер). С использованием алгоритма *ATRAC* удалось разместить запись длительностью звучания 74 мин на диске диаметром 64 мм и емкостью 140 Мбайт путем 5-кратного сжатия (5:1) аудиоданных по сравнению с обычными компакт-дисками (*CD*). Согласно утверждениям авторитетных экспертов и субъективным оценкам слушателей, потеря качества звучания практически неощутима.

Чуть позже в 1993 году этот алгоритм был использован в системе пространственного звучания *SDDS* для записи многоканального звука. В системе *SDDS* скорость цифрового потока составляет 292 кбит/с на канал.

В настоящее время известно несколько модификаций данного алгоритма (табл.2.10). Это свидетельствует о внимании фирмы к своему продукту, попытках постоянного его совершенствования с целью повышения качества компрессированных сигналов и снижения скорости цифрового потока, что необходимо также и для его продвижения в других возможных сферах применения. При этом более верхние версии остаются совместимыми с более ранними.

Следует все же отметить, что у самых первых экспериментальных версий *MD* использовалось 12-битовое нелинейное квантование коэффициентов преобразования с частотой дискретизации ЗС равной 32кГц, следовательно, ни о каком *Hi-Fi* качестве речи быть не могло. И только появление системы адаптивного кодирования *ATRAC*, использующего психоакустические аспекты при компрессии, позволило вывести качество записи *MD* на уровень, не уступающий *CD*, а в чем-то даже превосходящий его. Все же некоторым недостатком системы *ATRAC* является ее закрытость, т.е. по-

Таблица 2.10

Варианты алгоритма компрессии ATRAC

Версии алгоритма *ATRAC*, год появления на рынке	Скорость цифрового потока, кбит/с на канал
ATRAC-1; 1992 год	292
ATRAC-2; 1994 год	292
ATRAC-3; 1995 год	292
ATRAC-3.5; 1996 год	292
ATRAC-4; 1996 год	292
ATRAC-4.5 (только для MD-деки); 1996 год	292
*ATRAC*3 (для MDLP); 2000 год	132, 105,66
ATRAC DSP Type-R; 2001 год	292
ATRAC DSP Type-S; 2002 год	292
ATRAC3plus; 2003 год	256,64,48

дробные описания алгоритмов не опубликованы, защищены патентным законодательством, а стало быть, не доступны, по крайней мере, легальным образом, для использования сторонними разработчиками в своих продуктах.

Кодер системы кодирования *ATRAC*

Структурная схема кодера *ATRAC* представлена на рис. 2.82. На его

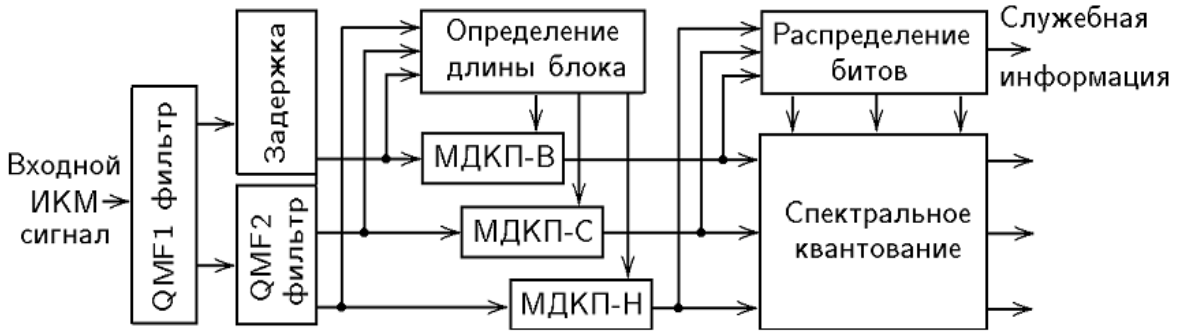


Рис. 2.82. Структурная схема кодера *ATRAC*

вход поступает ИКМ-сигнал с частотой дискретизации равной 44,1 кГц и разрешением 16 бит/отсчет, так что скорость цифрового потока составляет здесь 705,6 кбит/с. Входной сигнал обрабатывается квадратурным зеркальным фильтром (QMF1-фильтр), с помощью которого он разделяется на две одинаковые по ширине субполосные компоненты с полосами частот:

0…11,025 и 11…22,05 кГц. Далее один из этих субполосных сигналов (11,025…22,05 кГц) задерживается на время Δτ линией задержки ЛЗ (на интервал времени, необходимый для обработки одного из полученных сигналов в другом аналогичном фильтре), а другой (0…11,025 кГц) поступает на второй фильтр (QMF2-фильтр), где он также разделяется на две равные по полосе компоненты: 0..5,5125 и 5.5125…11,025 кГц. Итак, оба этих фильтра образуют банк фильтров, с помощью которого получаются три субполосных компоненты: 0…5,5125; 5,5125…11,025; 11,025…22,05 кГц исходного сигнала. Банк фильтров построен по древовидной структуре. Ее достоинством является отсутствие искажений, в местах стыковки субполосных сигналов, где имеет место их интерференции. Кроме того, на всех ступенях разделения и последующего синтеза (что необходимо в декодере) используются фильтры с одинаковым набором коэффициентов, это немаловажно с позиций их реализации.

Длина выборки составляет здесь 1024 отсчета звукового сигнала. После расфильтровки исходного ЗС имеем в полосе частот 11,025…22,05 кГц 512 отсчетов ЗС, а в двух других субполосах – соответственно по 256 отсчетов ЗС. Далее для уменьшения числа кодируемых элементов в каждой из этих субполос выполняется прямое модифицированное дискретное косинусное преобразование, обозначенное на рис. 2.82 соответственно как МДКП-В, МДКП-С и МДКП-Н, где буквы В, С, Н соответствуют верхним, средним и низким частотам. Таким образом, как и в алгоритме компрессии *MPEG Layer* 3 банк фильтров, включающий блок прямого ортогонального преобразования является гибридным.

Для уменьшения искажений, вызванных прямым (кодер) и обратным (декодер) ортогональным преобразованием, группы отсчетов ЗС предварительно взвешиваются оконными функциями (рис. 2.83). При этом используются два вида оконных функций – длинные (рис. 2.83,*а*, *Long*; их длина составляет 11,6 мс) и короткие (рис. 2.83,*б*, *Short*; их длина равна 1,45 мс в субполосе кодирования 11,025…22,05 кГц и 2,9 мс в субполосах кодирования 0..5,5125 и 5,5125…11,025 кГц). В отличие от кодера MPEG Layer 3 форма оконных функций выбрана таким образом, что не требуется так называемых «*окон перехода*» при переходе от длинных выборок к коротким и наоборот. При этом в каждой из субполос кодирования может использоваться одна из двух возможных их вариантов: в полосе 11,025…22,05 кГц длинные 11,6 мс и короткие 1,45 мс длины, а в полосах частот 0..5,5125 и 5.5125…11,025 кГц длинные 11,6 мс и короткие длиной в 2,9 мс. Кроме того, здесь принято также 50-ти процентное перекрытие по времени входных выборок звукового сигнала. После выполнения МДКП получаем в общей сложности 512 коэффициентов преобразования: 256 коэффициентов в полосе частот 11,025…22,05 кГц и по 128 коэффициентом МДКП в субполосах 0…5,5125 и 5,5125…11,025 кГц. Выбор длин оконных функции в субполосах кодирования определяется формой временной -

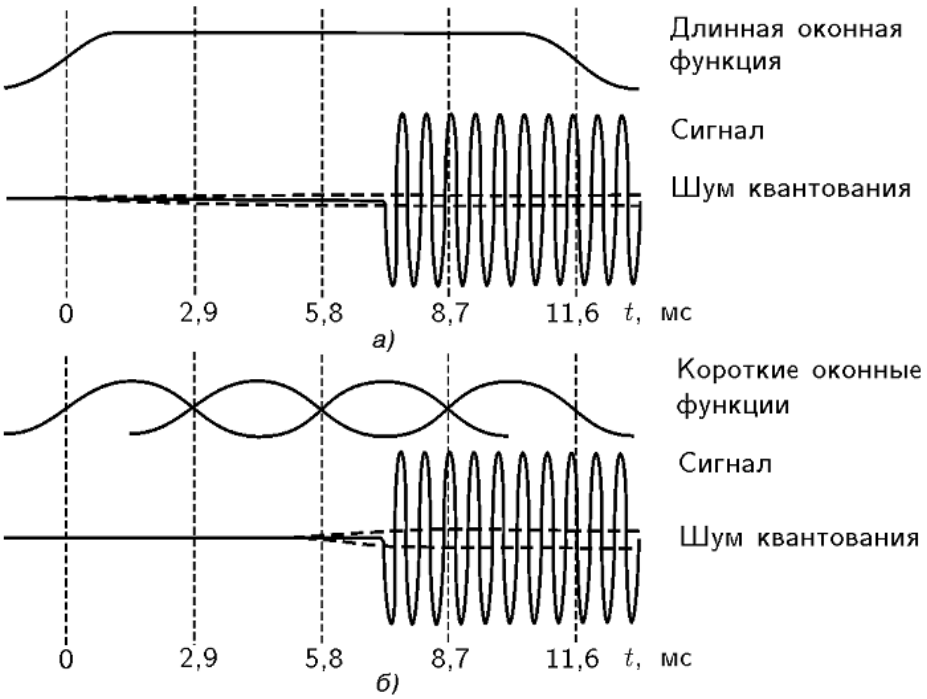


Рис. 2.83. Оконные функции при вычислении МДКП:
а – длинная оконная функция (16 мс для всех субполосных сигналов); *б* – короткие оконные функции (длина каждой из них составляет 2,9 мс в полосах частот 5,5…11 и 0….5,5 кГц и 1,45 мс в полосе частот 11…22 кГц)

функции сигнала. Для более или менее однородных выборок используются длинные оконные функции, а для неоднородных там, где имеют место резкие выбросы, используются короткие оконные функции. Этим учитывается динамика изменения сигнала внутри выборки. В первом случае мы имеем высокое разрешение по частоте, а во втором – по времени. Четкого критерия для перехода от коротких окон к длинным и наоборот в публикациях найти не удалось. Определение размера блока выполняется в блоке «*Block Size Decision*» кодера.

Заметим, что в других версиях алгоритма *ATRAC* возможно иное деление входного ЗС на субполосные сигналы. Например, в версии *ATRAC*3 *LP*2, использующей скорость передачи цифровых данных равной 132 кбит/с и обеспечивающей при этом (по мнению разработчиков) такое же качество, как и кодер *MPEG Layer* 3 при скорости цифрового потока равной 128 кбит/с, звуковой сигнал разделяется банком QMF-фильтров на четыре субполосные составляющие с полосам частот: 0…2,7562; 2,7652…5,5125; 5,5125…11,025 и 11,025…22,05 кГц. Кроме того, здесь на частотах выше 17,5 кГц дополнительно используется также процедура объединения сигналов стереопары. В версии же *ATRAC*3 *LP*4 при скорости передачи цифровых данных равной 66 кбит/с объединение сигналов стереопары при их кодировании выполняется уже на частотах выше 13,5 кГц. И наконец, в кодере *ATRAC3plus*, который используется в *HiMD*-плеерах,

применен банк фильтров, разделяющий входной сигнал до выполнения процедуры МДКП на 16 субполосных составляющих, благодаря чему удалось достичь скорости цифрового потока равной 64 кбит/с. Разработчики фирмы *Sony* поставили своей целью в кодере *ATRAC3plus* при скорости цифрового потока равной 64 кбит/с достичь качества алгоритма *MP*3, обеспечиваемого им при скорости цифрового потока равной 128 кбит/с. Так ли это на самом деле пока не совсем ясно.

Квантование коэффициентов МДКП выполняется с учетом психоакустики. При этом алгоритм компрессии *ATRAC* учитывает следующие свойства слуха и особенности восприятия звуковых сигналов при их обработке в слуховом анализаторе человека:

-кривые равной громкости, говорящие о том, что два достаточно узкополосных звуковых сигнала с одинаковым уровнем энергии, но с разными средними частотами, не будут восприниматься на слух равногромкими. Параметром каждой такой кривой равной громкости является уровень громкости, форма этих кривых зависит от уровня звука. В области меньшей чувствительности слуха искажения, вызванные квантованием кодируемых элементов, будут менее заметны на слух; при этом максимальная чувствительность слуха лежит в области 3…4 кГц; именно в этой области наибольшей чувствительности слуха малейшие изменения энергии сигнала, связанные с неточностью квантования сигнала будут заметны слушателям;

-абсолютный порог слышимости – минимальный уровень сигнала, еще воспринимаемый слухом в тишине, выражают в дБ; спектральные компоненты звукового сигнала, лежащие ниже абсолютного порога слышимости кодировать и передавать нет необходимости;

-обработка ЗС в слуховой системе человека осуществляется независимо в критических полосах (частотных группах) слуха; эти полосы имею разную ширину: около 100 Гц частотах ниже 500 Гц, выше этой частоты их ширина возрастает пропорционально частоте и составляет на самых верхних частотах более 3500 Гц (табл.2.11);

-маскировка одного звука в присутствии другого, можно говорить об одновременной и временной маскировке. В первом случае можно говорить о маскировке как внутри, так и вне критической полосы слуха, важно, что внутри критической полосы слуха и в сторону верхних частот она проявляется сильнее. При учете маскировки во временной области принято различать предмаскировку и постмаскировку, при этом предмаскировка ощутима на интервале времени 8…10 мс, а постмаскировка – на интервале времени 150…250 мс;

-тот факт, что в области частот до 5000 Гц критические полосы слуха являются достаточно узкими (их ширина не превышает 900 Гц) и в этой области (табл.2.11) находится около 18 критических полос слуха (две трети от их общего числа), свидетельствует о том, что в этой области частот ана-

Таблица 2.11

Границы и ширина полосы критических полос слуха, принятые в алгоритме компрессии *ATRAC*

Критические полосы слуха

Номер полосы	Частота, Гц			Номер полосы	Частота, Гц		
	Нижняя	Верхняя	Ширина полосы		Нижняя	Верхняя	Ширина полосаы
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

лиз сигнала слухом выполняется более точно. При этом он получает больше информации при анализе в этой области, а значит любые изменения сигнала, включая и появление искажений, будут наиболее заметны; квантовать элементы сигнала в этой области следует более точно; в то время как в верхней части спектра квантование сигнала может быть более грубым;

-величина шага квантования остается постоянной для группы кодируемых элементов, в каждую такую группу входят коэффициенты МДКП, лежащие в одной критической полосе слуха, следовательно, общее число таких групп равно числу критических полос слуха;

-на кодирование коэффициентов МДКП каждой критической полосы выделяется определенное число бит, эта величина к тому же зависит также и от выбранной скорости цифрового потока. Именно последняя определяет доступное для кодирования число бит. Из этого числа должны быть исключены биты заголовка (*Header*) и биты служебной информации (*Side Info*).

Все же точного упоминания о том, какое число полос психоакустического анализа здесь принято, в публикациях не указано, есть отдельные сведения о том, что их число равно 52-м, но вполне возможно, что в качестве них взяты критические полосы слуха, которых, как известно, 24.

Все эти перечисленные выше особенности слуха учтены при выборе процедуры распределения бит в кодере *ATRAC*. И еще одно важное замечание. При обработке малых по уровню сигналов применяется технология так

называемых *плавающих блоков*, в результате чего слабые сигналы обрабатываются с более высокой степенью разрешения. Суть этой технологии заключается в том, что слабоуровневые музыкальные фрагменты усиливаются и, как следствие, преобразование в цифру происходит более точно, а при воспроизведении искусственно "приподнятые" звуковые сигналы пересчитываются к исходному уровню. Описанный процесс тождественен преобразованию аналогового сигнала в системе шумоподавления фирмы *Dolby Lab*. Указанное преобразование существенно уменьшает искажения. Оно также почти на 20дБ расширяет динамический диапазон обрабатываемых сигналов.

До выполнения процедуры квантования в блоке «*Block Size Decision*» коэффициенты МДКП группируются в гранулы (блоки, группы), обозначаемые как *BFU* (рис. 2.84). Число коэффициентов МДКП в каждой такой

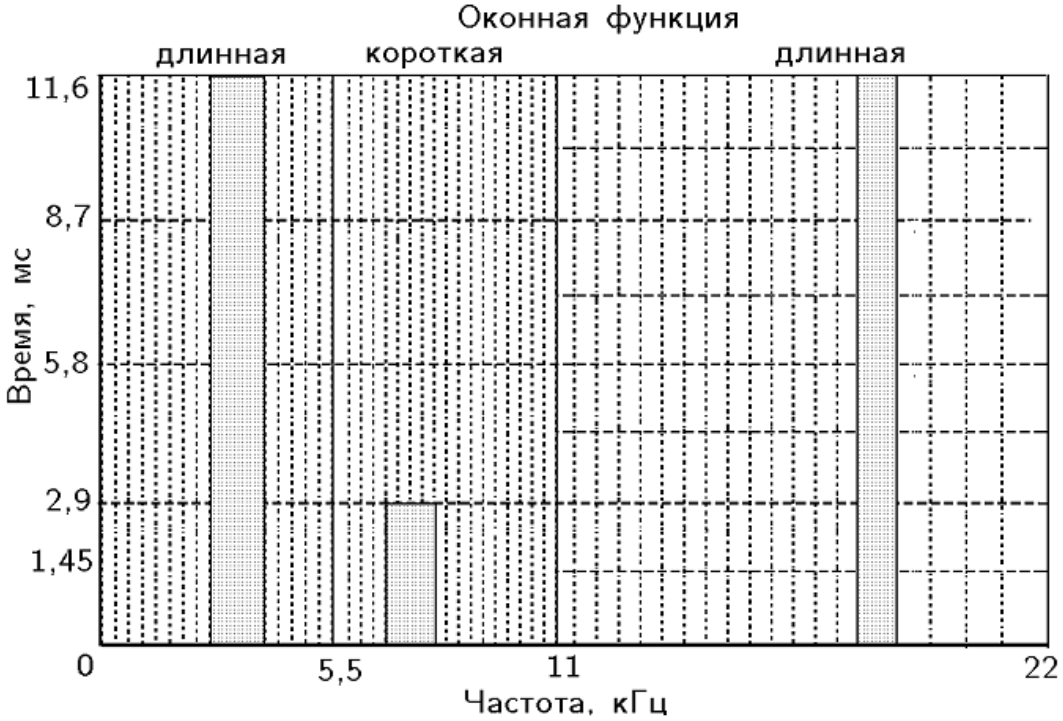


Рис. 2.84. К процедуре группирования коэффициентов МДКП перед их квантованием и кодированием

грануле определяется длиной оконной функции (рис. 2.83), длина таких блоков разная и выбирается адаптивно с помощью оконной функции. Она неодинакова в разных субполосах кодирования. Важно, что число блоков на нижних и средних частотах (полосы 0…5,5125 и 5,5125…11,025 кГц) существенно больше, чем на верхних (11,025…22,05). Это отражает свойства слуха.

Все доступные для кодирования коэффициентов МДКП биты распределяются между этими блоками BFU_k. Причем меньшее количество бит отводится на кодирование коэффициентов МДКП в той части спектра, где

чувствительность слуха ниже, а эта область соответствует более высоким частотам или, что тоже самое коэффициентам МДКП с более высокими значениями индексов. Внутри каждого бока BFU_k все коэффициенты МДКП квантуются с одинаковым значением шага квантования, величина этого шага меняется при переходе от одного такого блока к другому. Заметим, что шум квантования (рис. 2.83) равномерно распределен по каждому блоку и начальный его участок может быть не закрыт полезным сигналом, это явление названо *пред-эхом*. Это может привести к тому, что шум квантования станет заметен при прослушивании декодированного сигала. Переход к коротким блокам позволяет сделать заметность пред-эха меньшей за счет явления предмаскировки.

Для каждого блока BFU_k определяется масштабный коэффициент (нормирующий множитель), аналогично тому, как это делается, например, в *MPEG Layer* 2. Перед кодированием каждый коэффициент МДКП представляется в формате с плавающей запятой, аналогично тому, как это реализовано, например, алгоритме компрессии *Dolby AC*-3. Значения масштабных коэффициентов кодируются отдельно, при кодировании же мантисс учитываются перечисленные выше свойства слуха. Следовательно, для каждого блока BFU_k коэффициентов МДКП в цифровом потоке передаются следующая информация: длина оконной функции (она определяет число коэффициентов в блоке), кодовое слово масштабного коэффициента, длина кодового слова для каждого из коэффициентов МДКП в блоке и кодовые слова мантисс коэффициентов преобразования. Часть этой наиболее важной для правильного декодирования информации может быть дополнительно защищена.

Процедура квантования и кодирования здесь названа спектральной. Тем самым еще раз подчеркивается, что процедуре квантования и кодирования подвергаются не сами отсчеты ЗС, а соответствующе им коэффициенты МДКП.

Теперь более подробно рассмотрим процедуру распределения бит в алгоритме *ATRAC*. Прежде всего, доступное для кодирования количество бит делится между блоками BFU_k. Чем большее число бит из доступного их числа выделяется на кодирование блока, тем квантование входящих в него коэффициентов МДКП будет более точным, шумы квантования будут меньше, а длина кодовых слов коэффициентов МДКП соответственно больше, что очевидно. При малом числе выделенных бит картина будет обратной. Важно, что в алгоритме *ATRAC* процедура распределения бит между блоками жестко не задана, что необходимо для его дальнейшего совершенствования. Здесь возможно много разных способов от весьма простых до очень сложных. Однако, как утверждают разработчики, даже простой алгоритм распределения бит между блоками, если он учитывает психоакустику восприятия, может дать хорошие результаты. Используемое здесь

временное и частотное представление исходного ЗС перед его кодированием уже само по себе учитывает свойства слуха.

Один из предложенных способов в алгоритме *ATRAC* состоит в следующем. Общее взвешенное количество бит $b_{tot}(k)$, которое выделяется на кодирование информации *к*-того блока BFU_k, разделяется на две части фиксированную $b_{fix}(k)$ и переменную $b_{var}(k)$. При этом для каждого блока BFU_k соответственно имеем

$$b_{tot}(k) = \alpha b_{\text{var}} + (1-\alpha) b_{fix},$$

где α – индекс тональности, напомним, что он определяет близость компоненты сигнала (блока BFU_k) к чистому тону или белому шуму, его значение равно 1 для тона и 0 для шума; с его помощью учитывается маскировка внутри критической полосе слуха. Следовательно, пропорция между фиксированной и переменной часть выделенных бит есть в нашем случае величина переменная. Таким образом, для сигналов близких к чистым тонам используемое для их кодирования количество бит будет относиться только к переменной части, в другом противоположном случае, только к постоянной части. Иначе говоря, при кодировании шумоподобных блоков BFU_k биты, отнесенные к переменной части $b_{var}(k)$ вообще не будут использованы, и наоборот.

Выше приведенное уравнение никак не связано с установленной для кодера скоростью цифрового потока на его выходе. Но как только эта величина задается, то среднее доступное для кодирования коэффициентов МДКМП блока BFU_k число бит b_{off} легко может быть найдено. При этом если при вычислении выражения

$$b(k) = \{b_{tot}(k) - b_{0\,ff}\}_{\text{öåëî å÷èñëî}}$$

мы получаем отрицательное число, то биты для кодирования данного блока не выделяются. Иллюстрацией этого алгоритма является рис. 2.85.

Декодер системы кодирования *ATRAC*

Структурная схема декодера *ATRAC* представлена на рис. 2.86. При декодировании необходимы гораздо более простые преобразования: прежде всего цифровой поток демультиплексируется и затем с помощью служебной информации «*Side Info*» реконструируются в блоке спектральной реконструкции «*Spectral Reconstraction*» декодера в каждой субполосе ко-

эффициенты МДКП. Далее выполняется обратное модифицированное дискретное косинусное преобразование ОМДКП (блоки ОМСД-Н, ОМДКП-С,

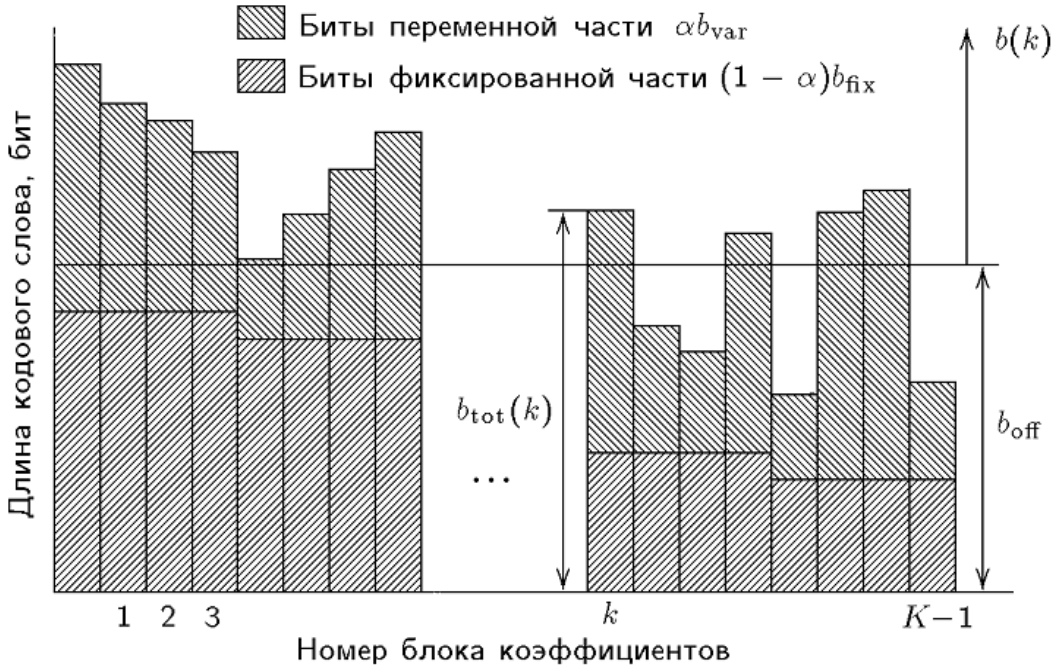


Рис.2.85. К процедуре распределения бит при кодировании блоков коэффициентов МДКП

ОМДКП-Н), где восстанавливаются субполосные составляющие исходного ЗС и затем в синтезирующих QMF-фильтрах они суммируются, образуя восстановленный звуковой сигнал.

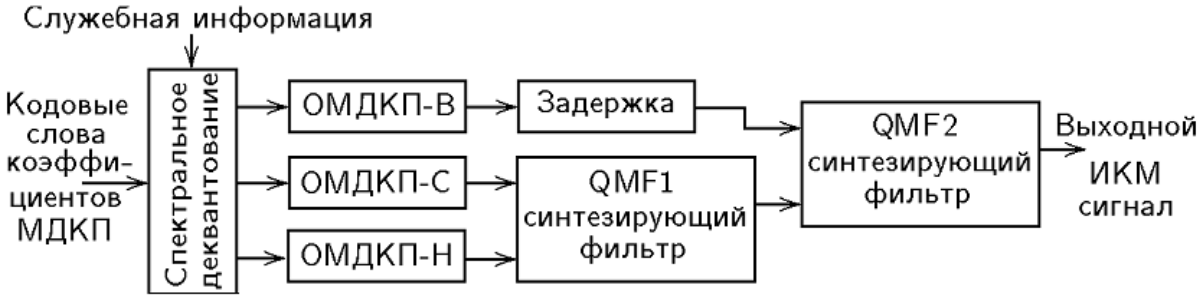


Рис. 2.86. Структурная схема декодера системы кодирования *ATRAC*

2.15. Качество алгоритмов компрессии цифровых аудиоданных

К настоящему времени уже известно достаточно большое число работ, посвященных исследованию качества звуковых сигналов, прошедших сжатие цифровых аудиоданных, оценке возникающих при этом артефактов, методам их минимизации.

Чаще всего эти исследования выполнены методом субъективно-статистических экспертиз, путем парного сравнения эталона и сигнала, подвергнутого компрессии, то есть прошедшего соответствующий кодек. При этом эталоном обычно выступает сигнал, поступающий на вход испытуемого кодека.

В качестве шкал оценки чаще всего выбираются пятибалльная и семибалльная шкалы (рекомендация *ITU-R BS*.562), представленные ниже.

5-ти балльная шкала оценки изменения качества:

0 баллов – незаметное различие,
1 балл – слабо заметное различие, но не раздражающее,
2 балла – заметное различие, слегка раздражающее,
3 балла – сильное различие, раздражающее,
4 балла – очень раздражающее различие.

В данной шкале приняты следующие градации ухудшения качества при парном сравнении отрывков звучаний:

0 – минус 1 балл (*Not annoying*),
минус 1 – минус 2 балла (*Slightly annoying*),
минус 2 – минус 3 балла (*Annoying*),
минус 3 – минус 4 балла (*Very annoying*);

7-ми балльная сравнительная шкала оценки качества:

+3 балла – намного лучше (*A much better than B*),
+2 балла – лучше (*A better than B*),
+1 балл – немного лучше (*A slightly than B*),
0 баллов – звучания равноценны (A same as B),
-1 балл – немного хуже *(A slightly worse than B*),
-2 балла – хуже (*A worse than B*),
-3 балла – намного хуже (*A much worse than B*)

При проведении таких субъективно-статистических экспертиз чаще всего используются отрывки звучаний разных жанров, взятые с компакт-диска *SQAM*, рекомендуемого для таких испытаний исследовательской группой *MPEG*.

Результаты экспертных оценок качества алгоритмов кодирования ЗС с компрессией цифровых аудиоданных, заслуживающие наибольшего доверия, представлены в табл. 2.12. Они получены для обычный двухканальной стереофонической системы. Экспертизы выполнены в соответствии с рекомендацией 562-3 «Субъективная оценка качества звука» и рекоменда-

циями *ITU-R BS*.1115 « *Low Bit Rate Audio Coding*», 1997, и *ITU-R BS*.1116-1 «*Methods for subijective assessment of small impairments in audio systems including multichannel sound systems»*, 1997. Здесь и ниже оценки даны в баллах.

Таблица 2.12

Результаты оценки качества кодеков с компрессией цифровых аудиоданных

Наименование алгоритма компрессии	Скорость цифрового потока на канал			
	96 кбит/с	128 кбит/с	160 кбит/с	192 кбит/с
MPEG-2 *AAC*	-1,15	-0,17	-	-
MPEG-1 *Layer* 3	-	-1,73	-	-
Dolby AC-3	-	-2,11	-1,04	-0,52
MPEG-1 *Layer* 2	-	-2,14	-1,75	-1,18

Наиболее детальные исследования для двухканальной системы были выполнены в 1998 году. Результаты этих исследований представлены на рис. 2.87. Здесь по вертикальной оси отложено различие в звучании сравниваемых пар отрывков (эталона, сигнал на входе кодека, и сигнала, прошедшего соответствующий кодек). На горизонтальной оси указаны жанры тестируемых отрывков реальных звучаний, взятых с компакт-диска *SQAM* группы *MPEG*. Каждая из представленных здесь кривых соответствует определенному алгоритму компрессии и значению установленной скорости

Таблица 2.13

Результаты оценки качества алгоритмов компрессии
(двухканальное воспроизведение)

Номер группы	Алгоритм компрессии	Скорость цифрового потока, кбит/с	Различие в звучании компрессированного сигнала в сравнении с эталоном, баллы
1	AC-3	128	–0,47
	AC-3	192	–0,52
2	PAC	160	–0,82
3	PAC	128	–1,03
	AC-3	160	–1,04
	AAC	96	–1,15
	Layer 2	192	–1,18
4	IT IS	192	–1,38
5	Layer 3	128	–1,73
	Layer 2	160	–1,75
	PAC	96	–1,83
	IT IS	160	–1,84
6	AC-3	128	–2,11
	Layer 2	128	–2,14
	IT IS	128	–2,21
7	PAC	64	–3,09
8	IT IS	96	–3,32

цифрового потока аудиоданных. Сравнительные данные для разных алгоритмов компрессии и влияние скорости цифрового потока на качество для этих же алгритмов представлено на рис. 2.88.

Для большего удобства сравнения все представленные выше результаты сведены в табл.2.13.

Ранжирование результатов оценок по степени деградации качества позволяет расположить алгоритмы с компрессией цифровых аудиоданных в следующем виде (табл. 2.14)

Таблица 2.14

Результаты ранжирования алгоритмов компрессии
по степени деградации качества (двухканальное воспроизведение)

Ранг	Алгоритм	Деградация качества, баллы
1	AAC	−0,47
2	PAC	−1,03
3	Layer 3 (MPEG-1)	−1,73
4	AC-3	−2,11
5	Layer 2 (MPEG-1)	−2,14
6	IT IS	−2,21

Результаты сравнения качества звучания кодеков (табл.2.12-2.14) свидетельствуют о том, что наиболее эффективным из стандартизованных является алгоритм компрессии *MPEG*-2 *ISO/IEC* 13818-7 *AAC*. Вертикальные линии на рис. 2.87 и 2.88 оценивают так называемый *доверительный интервал* с вероятностью попадания в него экспертопоказаний равной 0,95.

Проведение субъективно-статитических экспертиз (ССЭ) является очень дорогостоящим мероприятием. В последние годы достаточно часто подобные исследования выполняют, используя специальное программное обеспечение (ПО), позволяющее измерить качество цифрового сигнала, подвергнутого ранее компрессии. При этом все подобные измерения выполняют в полном соответствии с рекомендациями *ITU-R BS*.1116-1 и *BS*.1534-1. При этом результаты оценки также получают в баллах.

Для ряда кодеков, разработанных в России, подобные измерения выполнены А.С.Ивановым. Их можно разделить на три основных этапа. На первом этапе объективной оценке качества подвергалась программная модель кодека *MPEG*-1 *ISO/IEC* 11172-3 Layer 2, любезно предоставленная Ленинградским отраслевым научно-исследовательским институтом радио (ЛОНИИР). На втором этапе объективной оценке было подвергнуто конкретное устройство – кодек *CDQPRima*-230 фирмы *CCS*, предоставленный ЛОНИИС. На третьем этапе, полученные данные сравнивались с результатами субъективных прослушиваний, выполненных другими авторами. Оценки проводились для разных значений скорости цифрового потока аудиоданных.

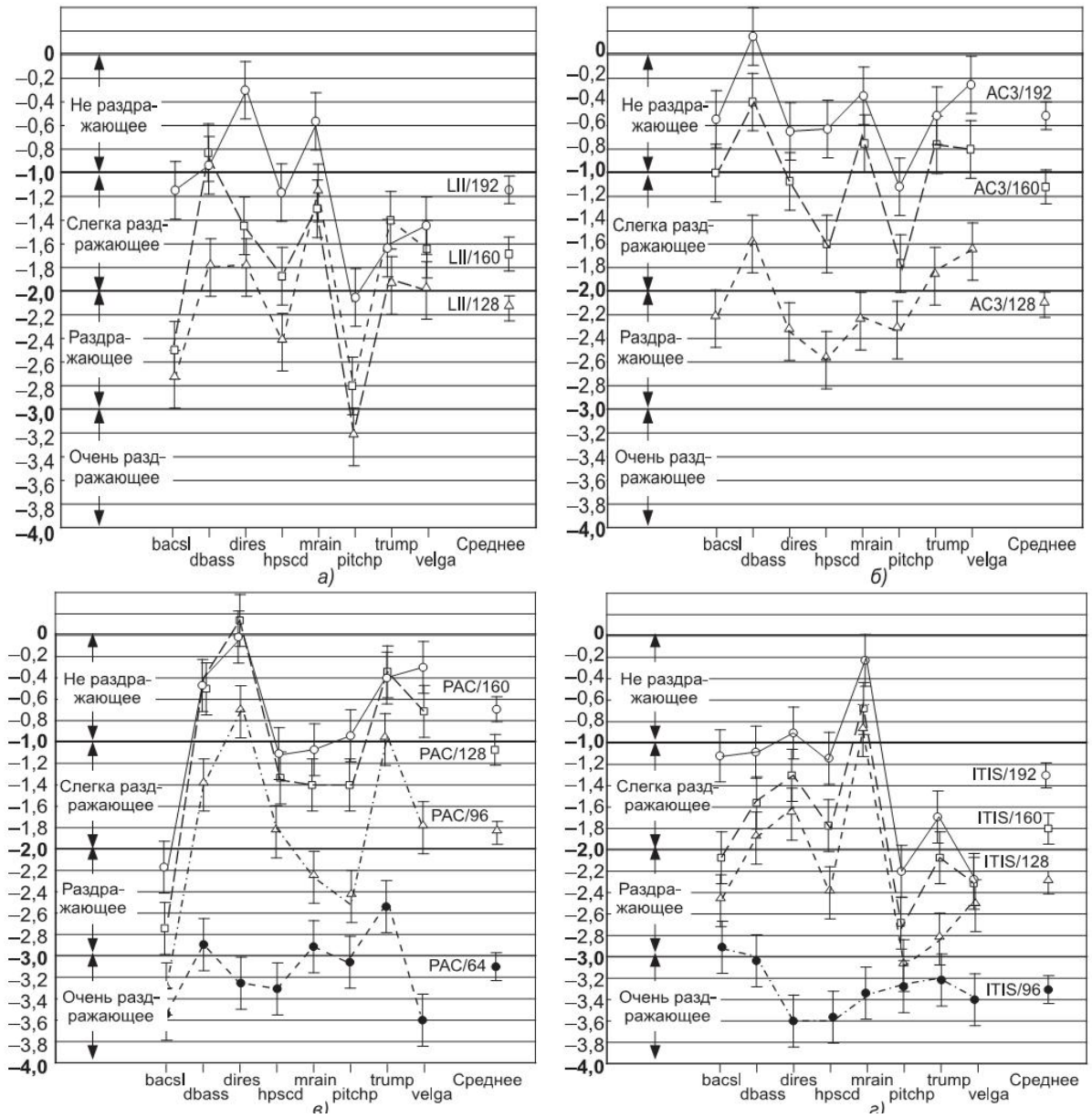


Рис. 2.87. Качество алгоритмов компрессии цифровых аудиоданных при двухканальном воспроизведении: а – кодек стандарта *MPEG*-1 *ISO/IEC* 11172-3 *Layer* 2 (LII) или *MPEG*-2 *ISO/IEC* 13818-3 *Layer* 2 (LII) при трех значениях скорости цифрового потока (128, 160 и 192 кбит/с); б – кодек стандарта *ATSC Dolby AC*-3 при трех значениях скорости цифрового потока (128, 160 и 192 кбит/с); в – кодек *PAC* (*Lucent Technologies*) для четырех значений скорости цифрового потока (64, 96, 128 и 160 кит/с); г – кодек *ITIS* также для четырех значений скорости цифрового потока данных (96, 128, 160 и 192 кбит/с)

В качестве испытательных сигналов здесь также использовались звуковые отрывки, рекомендованные группой *MPEG* для проведения ССЭ и позаимствованные с диска *EBU SQAM*. Все отрывки звуковых сигналов (ЗС) были представлены в формате *Windows* ИКМ, при этом частота дискретизации составляла 48 кГц и разрешение 16 бит/отсчет.

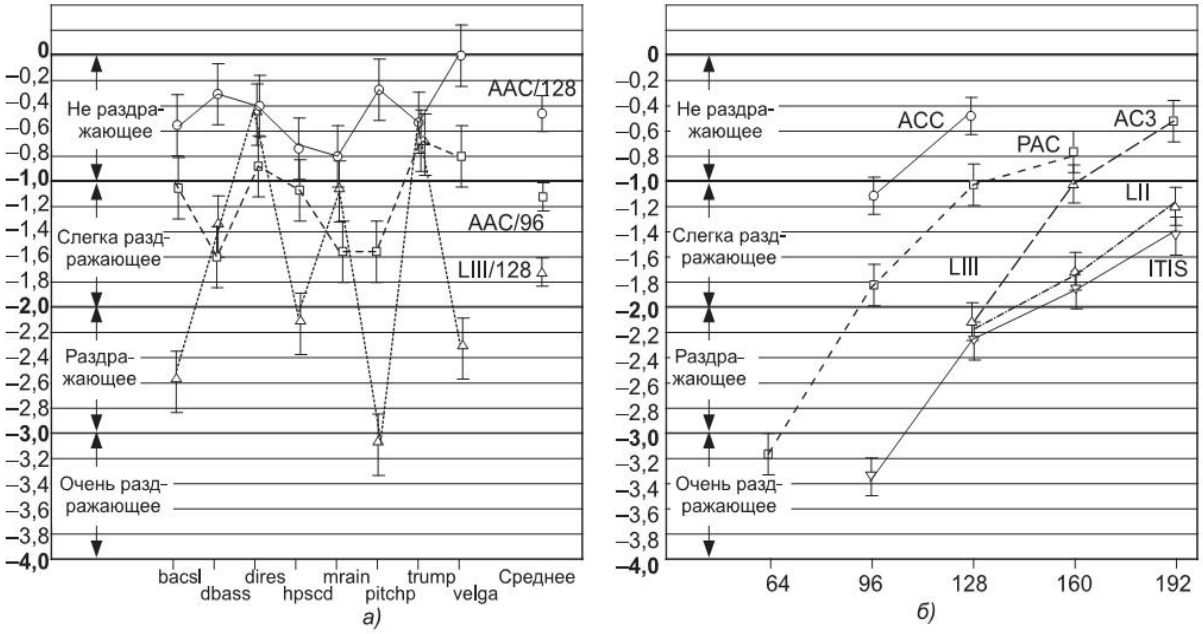


Рис. 2.88. Сравнительная оценка качества алгоритмов компрессии цифровых аудиоданных *Layer* 3 (LIII) и *AAC* для одних и тех же отрывков реальных звучаний *(а)* и зависимости изменения качества разных алгоритмов сжатия от установленного значения скорости цифрового потока (*б*).

При тестировании программной модели кодека *MPEG Layer* 2 (ЛОНИИР) скорости цифрового потока составляли 64, 128, 192, 256, 320 и 384 кбит/с для стереорежима работы. Для тестирования использовался набор испытательных сигналов, содержащий 31 отрывок ЗС. При испытаниях кодека *CDQPRima*-230 фирмы *CCS* (реализация кодеков *MPEG*-1 *Layer* 2 и *Layer* 3, скорости 64, 96, 128, 192 и 384 кбит/с, режим работы «стерео») использовалось три группы испытательных звуковых сигналов, по 25 сигналов в каждой группе, т.е. всего 75 сигналов.

Основные результаты тестирования кодеков представлены на рис. 2.89. Здесь приведены (для сравнения) также и соответствующие им субъективные оценки, заимствованные из опубликованных авторитетных зарубежных источников.

И, наконец, ниже даны последние сведения, представленные группой европейских исследователей на конгрессе общества *AES* в Вене в мае 2007 года (*EBU tests of multi-chanmel audio codecs*. *Convention Paper* 7052). Эти результаты получены при использовании кодеков с компрессией цифровых аудиоданных уже в системах *многоканальной стереофонии*, но также объективным путем с применением специального программного обеспечения *MUSHRA*. Расположение громкоговорителей соотвествует рекомендации *ITU-R BS*.775-2 «*Multichannel stereofonic sound system with and without accompanying picture*», *Juli* 2006. Результаты этих испытаний (рис. 2.90) представлены авторами в шкале *MUSHRA*, изменяющейся в пределах от 0 до 100 единиц с применением градаций оценки качества, представленной в

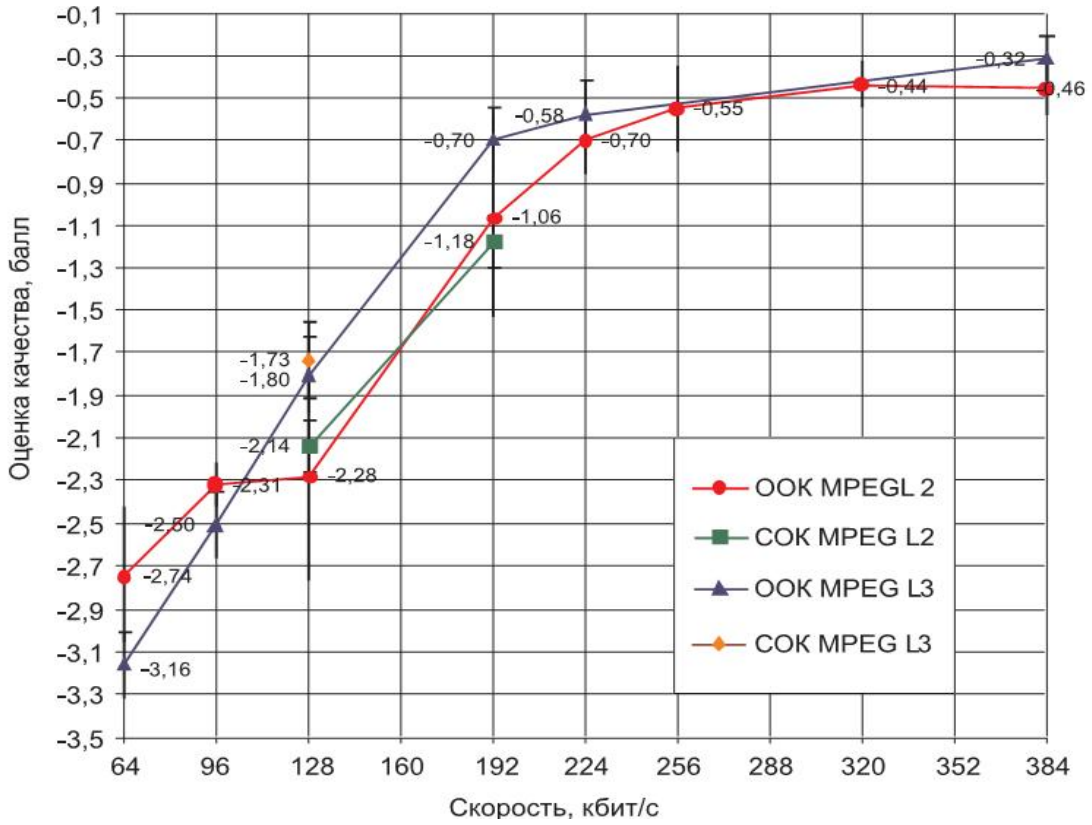


Рис.2.89. Деградация качества кодеков семейства *MPEG*-1 *Layer* 2 и *Layer* 3 (по данным А.С.Иванова)

(табл.2.15). В общей сложности было прослушано 700 отрывков разных жанров, каждый длительностью звучания около 30 с.

Таблица 2.15

Категории оценки качества звучания

Значения шкалы оценки	*Шкала субъективной оценки качества*
100-80	Превосходно (Excellenz)
80-60	Хорошо (Good)
60-40	Довольно (достаточно) хорошо (Fair)
40-20	(Poor)
20-0	Плохо (Bad)

На рис. 2.90 по вертикальной оси представлены типы испытанных кодеков, при этом приняты следующие сокращения: *DTS – Digital Theatre Sound*; *DD – Digital Dolby*; *DD+ – Dolby Digital Plus*; *WMA*10 – *Windows Media Audio* 10; *WMA9 – Windows Media Aud*io 9; *AAC – MPEG*-4 *Advanced Audio Coding*; *LII MPS – MPEG*-1 *Layer* 2 *with MPEG Surround*; *MP*3 *S – MPEG*-1 *Layer* 3 *with MPEG Surround*; *HEAACMPS – High Efficiency Advanced Audio Coding with MPEG Surround*; *PL*II *L*II – *Dolby Pro Logic* II *then MPEG Layer* 2; 3.5k – 3,5 *kHz low pass filtered anchor*; *Spatial Anchor – Re-*

duced image width anchor; Orig – оригинал. Число, расположенное рядом с аббревиатурой (рис. 2.90), есть ничто иное, как скорость цифрового потока аудиоданных, цифры даны в кбит/с за исключением системы *DTS*, где суммарное значение скорости цифрового потока равно 1,5 Мбит/с. По горизонтальной оси отложены градации оценки качества (табл.2.15). Самая нижняя строчка представляет собой оценку качества оригинала – исходного сигнала, поступающего на вход кодека. Это среднестатистические данные, усредненные для всех отрывков звучаний.

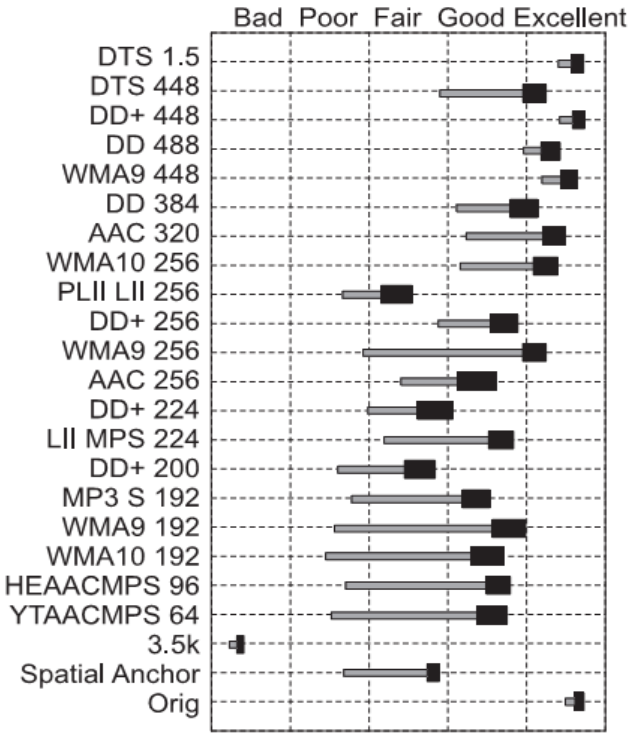


Рис. 2.90. Качество алгоритмов компрессии цифровых аудиоданных при многоканальном воспроизведении

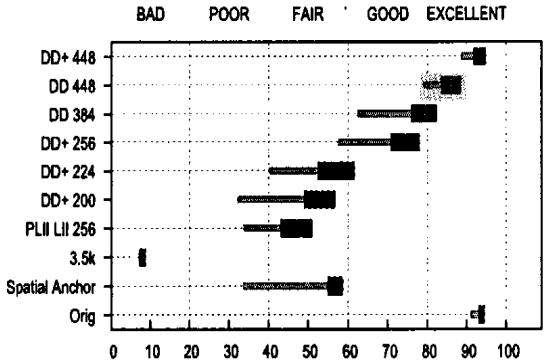
На рис. 2.91 показаны данные деградации качества алгоритмов в зависимости от скорости цифрового потока отдельно для систем *Dolby Digital* (*а*), Windows Media Audio (*б*) и *MPEG Audio* (*в*) также для многоканального воспроизведения.

Во всех представленных случаях (рис. 2.90 и 2.91) темная широкая часть – интервал, куда попадает 95% всех результатов оценок качества.

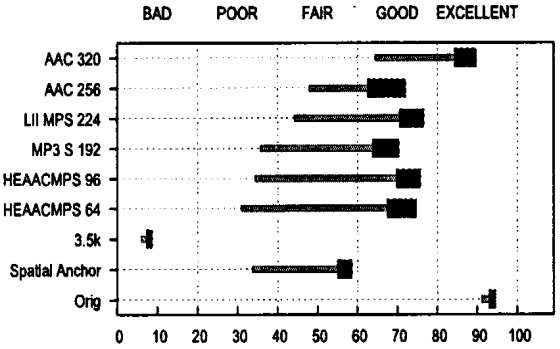
Из представленных выше данных следует, что превосходное качество при кодировании звуковых сигналов обеспечивается для:

-системы *DTS* при скорости цифрового потока 448 кбит/с,

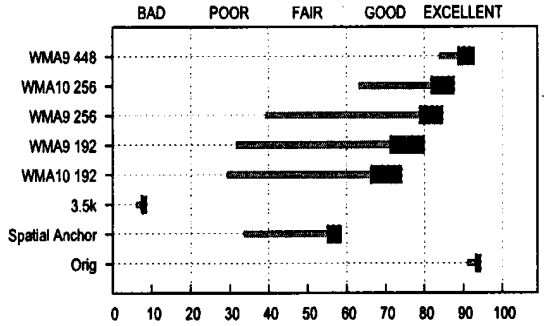
-системы *Dolby Digital* при скорости цифрового потока не менее 384 кбит/с,



а)



б)



в)

Рис. 2.91. Влияние установленного значения скорости цифрового потока на качество алгоритмов компрессии при многоканальном воспроизведении:
а – кодеки системы *Lolby*; б – кодеки *MPEG*; в – кодеки *Windows Media Audio*

-алгоритма *Windows Media audio* при скорости цифрового потока не менее 256 кбит/с,

-кодека *MPEG*-2 *Layer* 3 при скорости потока не менее 192 кбит/с.

Все же наиболее высокое качество обеспечивают алгоритмы сжатия *HEAACMPS*, при их применении в системах пространственного звучания скорость цифрового потока может быть уменьшена до 96-64 кбит/с (рис. 2.91,*б*).

Литература к разделу 1

1.1. Акустика: Учебник для вузов/в, Ю.А.Ковалгин, А.А.Фадеев, Ю.П.Щевьев; Под ред. профессора Ю.А.Ковалгина. - М.: Горячая линия – Телеком, 2009. – 660 с. (41,5 уч. изд. листа).

1.2. Алдошина И.А., Приттс Р. Музыкальная акустика. Учебник. - СПб.: Композитор, 2006.-720 с.

1.3. Вахитов Я.Ш. Слух и речь. Конспект лекций для студентов, обучающихся по специальности 0615 «Звукотехника». Л.: ЛИКИ, -1972.- 123 с.

1.4. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации. Перевод с немецкого под редакцией Б.Г.Белкина. М.: Связь, 1971. – 255 с.

1.5. Moore B., Glasberg B., Bauer T. A. A Model for the Prediction of Thresholds, Loudness and Partial Loudness/ Journal of Audio Engineering Society, vol. 45, № 4, April 1997. P.224-240

1.6. *Zwicker E., Fastl H.* Psychoacoustics: Facts and Models Second Ed. Verlag: Springer, 1999.

1.7. Zwiker E., Zwiker T. Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System/ Journal of Engineering Society, vol. 39, № 3, March 1991.

1.8. Zwiker E. Subdivision of the Audible Frequenvy Range into Critical Bands (Frequenzgruppen) / The Journal of Acoustical Society of America, Vol. 33, Number 2, February, 1961.- P.248.

1.9. Thiede T., Steinke G. Arbeitsweise und Eigenschaften von Verfahren zur Gehoerrichtigen Qualitaetsbewertung von Bitratenreduzierten Audiosignalen/Rundfunktechnische Mitteilungen, Nummer 3, 1994.-S. 102-114

1.10.Kapust R. Qualitaetsbeurteilung codierter Audiosignale mittels einer BARK-Transformation. Technische Fakultaet der Universitaet Erlangen-Nuernberg. Dissertation. Erlangen, 1993.

1.14. Plomp R. Aspects of Tone Sensation. London.: Academic Press, 1976.

1.15.E.Terhard. The SPINC function for Scaling of Frequency in Auditory Models/Acustica Vol. 77, 1992.

1.16.Rosisng T.D. The Science of Sound. New York.: Addision-Wesley Publ, 1982/

1.17. Therhard E. Akustische Kommunikation: Grundlagen mit Hoerbeispielen. Berlin, Heidelberg: Springer, 1998.

1.18. Bregman A.S. Auditory Science Analysis: The Perceptual Organization of Sound. Cambridge.: MIT Press, 1990.

1.19.Ковалгин Ю.А. и др. Акустические основы стереофонии / Ковалгин Ю.А., Борисенко А.В., Гензель Г.С. – М.: Связь, 1978.-336 с.

1.20. Fletcher H. Speech and hearing in Communication. New York: Van Nostrad, 1953.

1.21.ISO/IEC 11172-3: Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - part 3: Audio, 1993; 17

1.22. Colomes C., Lever M., Rault J.B., Dehery Y.-F., Faucon G., "A Perceptual Model Applied to Audio Bit-Rate Reduction", J. Audio Eng. Soc., vol.43(4), April 1995, pp.233-238; 9

1.23. Schroeder M.R., Atal B.S., Hall J.L., "Optimizing Digital Speech Coders by Exploting Masking Properties of the Human Ear", J.Acoust.Soc.Am.,vol.66(6), December 1979; 37

1.24. Mahieux Y., Petit J.P., "High-Quality Audio Transform Coding at 64 kbps", IEEE Trans. on Communications, vol.42(11), November 1994, pp.3010-3019; 29

1.25. ITU-R Document TG 10-2/3, Oct. 1991.

1.26. Brandenburg K., "Ein Beitrag zu den Verfahren und Qualitätsbeurteilung für hochwertige Musikcodierung", Erlangen-Nürnberg, Universität (Lehrstuhl für Technische Elektronik), Dissertation, 1989.

1.27. Zwicker E, Feldtkeller R. "Das Ohr als Nachrichtenempfänger.", Stuttgard: S.Hirzel Verlag, 1967; 53

1.28. Solbach, L.An Architecture for Robust Partial Tracking and Onset Localization in Single Channel Audio Signal Mixes. Dissertation, 1998, http://www.ti6.tu-harburg.de/~ti6ls/diss.

1.29. Brian C.J.Moore. Masking in the human auditory system // Collected Papers on Digital Audio Bit-Rate Reduction (Journal of the audio engineering society). – 1996. – P. 9-19.

1.30. Jesteadt W., Bacon S. P., Lehman J. R. Forward masking as a function of frequency, masker level, and signal delay // J. Acoust. Soc. Am. 1982. Vol. 71. P. 950–962.

1.31. Fastl H. Temporal masking effects: I. Broad band noise masker // Acustica. 1976. Vol. 35, S.287-302.

1.32. Fastl H. Temporal masking effects: II. Critical band noise masker // Acustica . 1977. Vol. 36. S. 317–331.

1.33. Fastl H. Temporal masking effects: III. Pure tone masker // Acustica. 1979. Vol. 43, S. 282-294.

1.34. Plomp R. The rate of decay of auditory sensation // J. Acoust. Soc. Am. 1964. Vol. 36. P. 277–282.

1.35. Stein H. *J.* Das Absinken der Mitho¨rschwelle nach dem Abschalten von weißem Rauschen // Acustica. 1960. Vol. 10. S. 116–119.

1.36. Moore B.C.J., Glasberg B. R. Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise // J. Acoust. Soc. Am. 1983. Vol. 73. P. 1249–1259.

1.37. Zwicker E., Fastl H. Zur Abhängigkeit der Nachverdeckung von der Störimpulsdauer Acustica. 1972. Vol. 26. S. 78-82.

1.38. Zwicker E. Dependence of post-masking on masker duration and its relation to temporal effects in loudness // J. Acoust. Soc. Am. 1984. Vol. 75. P. 219–223.

1.39. Widin G. P., Viemeister *N.F.* Intensive and temporal effects in pure-tone forward masking // J. Acoust. Soc. Am. 1979. Vol. 66. S. 388-395.

1.40. Duifhuis H. *J.* Consequences of peripheral frequency selectivity for no simultaneous masking // Acoust. Soc. Am. 1973. Vol. 54(6). Dec. P. 1471-1488.

1.41. Nelson D.A., Freyman R.L. Temporal Resolution in Sensor neural Hearing-Impaired Listeners // J. Acoust. Soc.Am. 1987. Vol. 81. P. 709-720.

1.42. Moore B.C.J., Glasberg B.R., Plack C.J., Biswas A.K. The shape of the ear's temporal window // J. Acoust. Soc. Am. 1988. Vol. 83. P. 1102–1116.

1.43. Plack C.J., Moore B.C.J. Temporal window shape as a function of frequency and level // J. Acoust. Soc. Am. 1990. Vol. 87. P. 2178–2187.

1.44. Plack C. J., Oxenham A. J. Basilar-membrane nonlinearity and the growth of forward masking // J. Acoust. Soc. Am. 1998. Vol. 103. № 3. P.1598-608.

1.45. Hawksford M.O.J., Hollier, M. P. (1993) "Characterizationof Communications-SystemsUsing a Speechlike Test Stimulus" J. Audio Eng. Soc., 41, No. 12,

1.46. Widin, G. P., Viemeister, N.F.(1979) "Intensive and temporal effects in pure-tone

forward masking” J.Acoust.Soc.Am., 66, 388-395

1.47. ITU-R Recommendation BS.1387. (1998). “Method for Objective Measurements of Perceived Audio Quality”.

1.48. Terhardt, E. (1979). “Calculating virtual pitch”. Hearing Research, 1, p.155-182.

1.49. Meddis, R., O'Mard, L.P.(2005) “A computer model of the auditory-nerve response to forward-masking stimuli” J.Acoust.Soc.Am 117(6), 3787-3798

1.50. Moore, B.C. (1978). “Psychophysical tuning curves measured in simultaneous and forward masking.” J Acoust Soc Am. Feb;63(2):524-32.

1.51. Houtgast, T.(1972) “Psychophysical Evidence for Lateral Inhibition in Hearing” J.Acoust.Soc.Am 51(6B)

1.52. Альтман Я.А. Локализация звука (нейрофизиологические механизмы). Л.: Наука, 1972.-214 с.

1.53. Schenkel K. Ueber die Abhaengigkeit der Mithoerschwellen von der interauralen Phasenlage des Testschals. – Acustica, vol. 4? 1964, S.337-346.

1.54. Schenkel K. Accumulation theory of binaural masked thresholds. –J.Acoust. Sos. Amer., vol. 41, N 1, 1967, p.20-30

1.55. Durlach N.J. Equalization and cancellation theory of binaural masking-level differences. - J. Acoust. Sos. Amer., vol. 35, 1963, p.1206-1218

1.56. Durlach N.J. On the application of the EC-model to interaural jund`s. - J. Acoust. Sos. Amer., vol. 40, N 6, 1966, p.162-181.

1.57. G.Тейле (Theile G. Zur Theorie der Optimalen Wiedergabe von Stereofonen Signalen ueber Lautsprecher und Korphoerer// Rundfunktechnische Mitteilungen. - 1981, J.25, Heft 4.- S.155-170

1.58. Ланге Ф. Корреляционная электроника. Л.: Судпромгиз, 1963.-447 с.

1.59.Cherry E.C., Sayers M.A. Human cross-correlator./ J. Acoust. Sos. Amer., vol. 28, N 5, 1956, p.889-895. Госэнергоиздат,1954.-524 с.

1.60. Ковалгин Ю.А. Стереофония. -М.: Радио и связь, 1989.-272 с.

1.61. Блауэрт Й. Пространственный слух: Пер. с нем. – М.: Энергия, 1979.-224 с.

1.62. Blauert J. Raeumlichen Hoeren. Nachschrift. Neue Ergebnisse und Trends seit 1972. – S.Hirzel Verlag Stuttgart, 1985.-119 S.

1.63. G. Kendall & W. Martens, «Simulating the cues of spatial hearing in natural environments» //Proceedings of the 1984 International Computer Music Conference.

1.64. A.W. Mills, «Auditory Localization» //In.: J.V. Tobias, ed., Foundations of Modem Auditory Theory. - Academic Press, 1972, vol. 2, p.337.

1.65. Schubert E. Some preliminary experiments on binaural time delay and intelligibility//J. Acoust. Soc. Amer., vol. 28, №5, 1956. p.456-464

Литература к разделу 2

2.1. Pohlman, K.C. Principles of Digital Audio, 5rd Ed. McGraw-Hill, 2005.-860 с.

2.2. Watkinson, J.R. The Art of Digital Audio. 2nd Ed.. Boston, MA: Focal Press, 1994.

2.3. Watkinson, J.R. Coding for Digital Recording. Boston, MA: Focal Press, 1990.

2.4. Stuart J.R. Coding for High-Resolution Audio Systems. J.Audio Eng.Soc. v. 52, N3, 2004, march, pp.117-144.

2.5. Электроакустика и звуковое вещание: Учебное пособие для вузов / И.А. Алдошина, Э.И.Вологдин, А.П. Ефимов и др.; Под ред. Ю.А.Ковалгина. -М.: Горячая линия - Телеком, Радио и связь, 2007.- 872 с.

2.6. Стереофоническое радиовещание и звукозапись: Учебное пособие для вузов/ Ю.А.Ковалгин, Э.И.Вологдин, Л.Н.Кацнельсон; Под ред. профессора Ю.А.Ковалгина.- М.: Горячая линия – Телеком, 2007, - 720 с.

2.7. Ковалгин Ю.А., Вологдин Э.И. Цифровое кодирование звуковых сигналов: Учебное пособие.- СПб.: КОРОНА-принт, 2004, - 240 с.

2.8. Попов О.Б.Б., Рихтер С.Г. Цифровая обработка сигналов в трактах звукового вещания: Учебное пособие для вузов. – М.: Горячая линия – телеком, 2007. – 341 с.

2.9. Оппенгейм А., Шафер Р. Цифровая обработка сигналов. – М.: Техносфера, 2006. – 856 с.

2.10. Сэломон Д. Сжатие данных изображений и звука. – М.: Техносфера, 2004. – 368 с.

2.11. Щербина В.И. Цифровая звукозапись. М.: Радио и связь, 1989.

2.12. ISO/IEC 14496-3, Information Technology - Coding of audio-visual objects - Part 3: Audio",1999. Super Audio CD Format.

2.13. ISO/IEC 13818-7. Information Technology -Generic coding of moving pictures and associated audio - Part 7: Advanced Audio Coding", 1997. Super Audio CD Player SCD-1 Technology.

2.14. International Standard ISO/IEC 11172-3. Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s- Part 3: Audio, 1993-08-01.

2.15. International Standard ISO/IEC 13818-3. Information technology-Generic coding of moving pictures and associated audio information. Part 3:Audio, 1995-05-15.

2.16. International Standard ISO/IEC 13818-7. Information technology-Generic coding of pictures and associated audio information. Part 7: Advanced Audio Coding (AAC), 1997 (E).

2.17. ISO/IEC FCD 14496-3 Subpart 1. Information Technology-Very Lov Bitrate Audio-Visual Coding. Part 3: Audio, 1998-05-10 (ISO/JTC 1/SC 29, N2203).

2.17. Digital Audio Compression Standard (AC-3). Doc.A/52, 1995-12-20.

2.18. Ковалгин Ю.А. Алгоритмы компрессии цифровых аудиоданных// Системы и средства связи, телевидения и радиовещания, номер 3, 2000. –с. 17-29.

2.19. Ятагама Гамаге Д.П. Повышение эффективности алгоритмов компрессии цифровых данных при кодировании сигналов стереопары. Автореферат диссертации на соискание ученой степени кандидата технических наук, СПб.: СПбГУТ 2005. - 16 с.

2.20. Зырянов М.В. Повышение эффективности алгоритмов компрессии цифровых аудиоданных на основе учета временной маскировки. Автореферат диссертации на соискание ученой степени кандидата технических наук, СПб.: СПбГУТ, 2007. – 16 с.

2.21. Захаренко А.В. Учет временных свойств слуха при сокращении психофизической избыточности звукового сигнала//Труды учебных заведений связи, СПб.: СПбГУТ, № 172, 2005.

2.22. Johnston J., Transform Coding of Audio Signals Using Perceptual Noise Criteria // IEEE J. Sel. Areas in Comm., pp. 314-323, Feb. 1988.

2.23. Kahrs M., Brandenburg K. Applications of Digital Signal Processing to Audio and Acoustics. – Kluwer Academic Publishers. New York, Boston, Dordrecht, London, Moscow. -535 p.

2.24. Spanias F., Painter A., Atti v. Audio Signal Processing and Coding. Wiley, 2007. -459 p.

2.25. Bosi M., Goldberg R. E. Introduction to Digital Audio Coding and Standards. – Springer, 2003. - 458 p.

2.26. Кацнельсон Л.Н. Системы цифрового радиовещания DAB, DMB и DAB+; Часть 1: учебное пособие; ГОУВПО СПбГУТ, 2009.- 100 с.

2.27. Кацнельсон Л.Н. Системы цифрового радиовещания DAB, DMB и DAB+; Часть 2: учебное пособие; ГОУВПО СПбГУТ, 2009.- 64 с.

2.28. Кацнельсон Л.Н. Системы цифрового радиовещания DAB, DMB и DAB+; Часть 3: учебное пособие; ГОУВПО СПбГУТ, 2009.- 68 с.

2.29. Кацнельсон Л.Н. Система цифрового радиовещания DRM. – СПб.: Изд.-во «Линк», 2010. – 76 с.

2.30. Шелухин О.И., Лукъянцев Н.Ф.Цифровая обработка и передача речи/Под ред О.И.Шелухина. – М.: Радио и связь, 2000.-456 с.

2.31. Документ ETSI TS 101 980 V1.1.1 (2001-09). Digital Radio Mondiale (DRM); System Specification.

2.32. Акустика:Учебник для вузов/ Ш.Я.Вахитов, Ю.А.Ковалгин, А.А.Фадеев, Ю.П.Щевьев; Под ред. Профессора Ю.А.Ковалгина. – М.: Горячая линия –Телеком, 2009.- 660 с.

2.33. Лемешко Б.Ю., Лемешко С.Б. Сравнительный анализ критериев проверки отклонения распределения от нормального закона // Метрология. 2005. №2.

2.34. R.Coifman, Y.Meyer, S.Quake, M.V.Wickerhauser, “Signal Processing and Compression with Wavelet Packet,” in Num. Alg. Res. Group, New Haven, CT: Yale University, 1990.

2.35. K.Hamdy, Low Bit Rate High Quality Audio Coding with Combined Harmonic and Wavelet Representations, in Proc. Int. Conf. Acous., Speech and Sig. Proc. (ICASSP-96), pp.1045-1048, May 1996.

2.36. ITU-R BS.1116.

2.37. ITU-R BS.562-3.

2.38. J.Princen and J.D.Johnston, Audio Coding with Signal Adaptive Filterbanks, in Proc. ICASSP-95, pp.3071 – 3074, May 1995.

2.39. D.Sinha and A.Tewfik, Low bit rate transparent audio compression using adapted wavelets, IEEE Trans. Signal Processing, vol.41, no.12, pp.3463 – 3479, December 1993.

2.40. A.Tewfik and M.Ali, Enhanced Wavelet Based Audio Coder, in Conf. Rec. of the 27th Asilomar Conf. on Sig. Sys., and Comp., pp.896-900, Nov 1993.

2.41. Breebaart J. et al. Parametric Coding of Stereo Audio//EURASIP Jornal on Applied Signal Processing, 2005. - № 9, pp. 1305-1322.

2.42. Breebaart J., Faller C. Spatial Audio Processing MPEG Surround and Other Applications. – John Wiley & Sons, Ltd. – 2007, 222 p.

2.43. Schuijers E., Breebaart J., Purnhagen H., Engdegard J. Low complexity parametric stereo coding. Proc. 116th AES convention, Berlin, Germany, Preprint 6073.

2.44. ISO/IEC 23003-1:2007, “Information Technology— MPEG Audio Technologies—Part 1: MPEG Surround,” International Standards Organization, Geneva, Switzerland (2007).

2.45. ISO/IEC 23003-1:2007/Cor.1:2008, “Information Technology—MPEG Audio Technologies—Part 1: MPEG Surround, TECHNICAL CORRIGENDUM 1,” International Standards Organization, Geneva, Switzerland (2008).

2.46. J. Herre, “From Joint Stereo to Spatial Audio Coding— Recent Progress and Standardization,” presented at the 7th Int. Conf. on Digital Audio Effects (DAFX04) (Naples, Italy, 2004 Oct.).

2.47. H. Purnhagen, “Low Complexity Parametric Stereo Coding in MPEG-4,” presented at the 7th Int. Conf. on Audio Effects (DAFX-04) (Naples, Italy, 2004 Oct.).

2.48.E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård, “Low-Complexity Parametric Stereo Coding,” presented at the 116th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 52, p. 800 (2004 July/Aug.), convention paper 6073.

2.48. C. Faller and F. Baumgarte, “Efficient Representation of Spatial Audio Using Perceptual Parameterization,” presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (New Paltz, NY, 2001).

2.49. C. Faller and F. Baumgarte, “Binaural Cue Coding— Part II: Schemes and Applications,” *IEEE* Trans. Speech Audio Process*.,* vol. 11 (2003 Nov.).

2.50. C. Faller, “Coding of Spatial Audio Compatible with Different Playback Formats,” presented at the 117th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts)*,* vol. 53, p. 81 (2005 Jan./Feb.), convention paper 6187.

2.51. R. Dressler, “Dolby Surround Prologic Decoder— Principles of Operation,” Dolby Publi.,

http://www.dolby.com/assets/pdf/tech_library/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf.

2.52. D. Griesinger, “Multichannel Matrix Decoders for Two-Eared Listeners,” presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 44, p. 1168 (1996 Dec.), preprint 4402.

2.53. J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Holzer, K. Linzmeier, C. Spenger, and P. Kroon, “Spatial Audio Coding: Next-Generation Efficient and Compatible Coding of Multichannel Audio,” presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 53, p. 81 (2005 Jan./Feb.), convention paper 6186.

2.54. ISO/IEC JTC1/SC29/WG11 (MPEG), “Call for Proposals on Spatial Audio Coding,” Doc. N6455, Munich, Germany (2004).

2.55. ISO/IEC JTC1/SC29/WG11 (MPEG), “Report on Spatial Audio Coding RM0 Selection Tests,” Doc. N6813, Palma de Mallorca, Spain (2004).

2.56. J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. Kjo¨rling, E. Schuijers, J. Hilpert, and F. Myburg,“The Reference Model Architecture for MPEG Spatial Audio Coding,” presented at the 118th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts)*,* vol. 53, pp. 693, 694 (2005 July/Aug.), convention paper 6447.

2.57. ISO/IEC JTC1/SC29/WG11 (MPEG), “Report on MPEG Spatial Audio Coding RM0 Listening Tests,” Doc.N7138, Busan, Korea (2005); available at http://www. chiariglione.org/mpeg/working_documents/mpeg-d/sac/ RM0-listening-tests.zip.

2.58. J. Breebaart, J. Herre, C. Faller, J. Roden, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K.Kjorling, and W. Oomen, “MPEG Spatial Audio Coding/ MPEG Surround: Overview and Current Status,” presented at the 119th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts*),* vol. 53, p. 1228 (2005 Dec.), convention paper 6599.

2.59. J. Breebaart, J. Herre, L. Villemoes, Craig Jin, K. Kjo¨rling, J. Plogsties, and J. Koppens: “Multi-ChannelGoes Mobile: MPEG Surround Binaural Rendering,” presented at the 29th AES Int. Conf. (Seoul, Korea, 2006). [18] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Purnhagen, and K. Kjo¨rling, “MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding,” presented at the 28th Int. Conf. (Piteå, Sweden, 2006).

2.60. B. R. Glasberg and B. C. J. Moore, “Derivation of Auditory Filter Shapes from Notched-Noise Data,” Hear. Research*,* vol. 47, pp. 103–138 (1990).

2.61. J. Breebaart, S. van de Par, and A. Kohlrausch, “Binaural Processing Model Based on Contralateral Inhibition— I. Model Setup,” J. Acoust. Soc. Am., vol. 110, pp. 1074–1088 (2001).

2.62. J. Princen, A. Johnson, and A. Bradley, “Subband/ Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation,” in Proc. IEEE ICASSP (1987), pp. 2161–2164.

2.63. M. Dietz, L. Liljeryd, K. Kjo¨rling, and O. Kunz, “Spectral Band Replication—A Novel Approach in Audio Coding,” presented at the 112th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 50, pp. 509, 510 (2002, June), convention paper 5553.

Ковалгин Юрий Алексеевич

ПСИХОАКУСТИКА
И КОМПРЕССИЯ ЦИФРОВЫХ АУДИОДАННЫХ

Издано в авторской редакции

План издания научной литературы 2012 г., п. 11

Подписано к печати 14.11.2012
Объем 18,75 усл.-печ. л. Тираж 500 экз. Заказ 244

Издательство СПбГУТ. 191186 СПб., наб. р. Мойки, 61
Отпечатано в мини-типографии изд-ва «Знакъ».
191011 СПб., Невский пр., 32–34