

Восприятие и цифровое представление звуковой информации

Введение

- Виды аппаратной и программной реализации систем цифровой обработки речевой и звуковой информации определяются их исходными характеристиками, особенностями слухового восприятия и требованиями к качеству воспроизведения
- Речевая информация, образующая свойственные используемому языку фонетические комбинации и формирующая те или иные смысловые элементы, по своим физическим параметрам принципиально отличается от звуковой информации, содержащей сочетание голосовых данных с музыкальным сопровождением, особенности и отличия друг от друга речевой и звуковой информации используются при их цифровой обработке и сжатии
- Основную информацию о звуковых колебаниях человек получает в области частот примерно до 4 кГц, именно эти частоты задают разборчивость и ясность аудиоинформации
- Спектральный состав речи занимает полосу частот примерно от 50 до 7000-10000 Гц
- В аналоговой телефонии используется полоса частот 0,3-3,4 кГц, что ухудшает восприятие ряда звуков (например, шипящих), но практически не отражается на разборчивости речи

- В цифровой телефонии отсчеты аналоговой речи приходится брать согласно теореме Котельникова с частотой 8 кГц
- Разрядность аналого-цифрового преобразования для речи – 8 или 16 бит на отсчет
- Идея преобразовывать в цифровой вид не сам речевой сигнал, а его параметры (количество переходов через ноль, спектральные характеристики и др.), чтобы затем по этим параметрам выбирать модель голосового тракта и синтезировать исходный сигнал, лежит в основе синтезирующих кодеков или вокодеров
- Принцип работы гибридных кодеков основан на модели кодирования с использованием линейного предсказания и алгебраической кодовой книги, при этом производится анализ речевого сигнала и выделяются параметры модели (коэффициенты системы линейного предсказания, индексы и коэффициенты усиления в адаптивной и фиксированной кодовых книгах), далее эти параметры кодируются и передаются в канал
- Слуховой аппарат человека различает частотные составляющие звука приблизительно в пределах от 30 Гц до 20 кГц; верхняя граница может несколько отличаться в зависимости от возраста человека, условий воспроизведения информации и др.

- Используемые для звука частоты дискретизации – 32, 44,1, 48, 96 кГц; при этом цифровой поток продискретизированного сигнала может изменяться от $32000 \cdot 16 / 1024 = 500$ кбит/с до $96000 \cdot 24 \cdot 6 / (1024 \cdot 1024) = 4,4$ Мбит/с и даже $192000 \cdot 32 \cdot 24 / (1024 \cdot 1024) = 140,6$ Мбит/с

Почему 44100?

- Частота 44.1 кГц возникла в конце 1970-х, благодаря PCM адаптерам, которые записывали звук на видеокассеты (U-Matic), в частности Sony PCM-1600 (1979) и последующим моделям серии. Позже это стало основой для CD-DA, описанного в стандарте Red Book (1980). В дальнейшем это значение частоты также было включено в другие стандарты 90-х/2000-х годов, вроде DVD, HDMI. Данная частота обычно используется при кодировании в MP3 (и другие потребительские форматы аудио) звука, извлеченного из Audio CD.
- Частота дискретизации была выбрана в ходе дебатов между разработчиками, в особенности Sony и Philips, а также благодаря компании Sony, которая в результате активного использования этой частоты практически сделала её стандартом де-факто. Само собой, выбор имел определенное техническое обоснование.
- Слышимый диапазон для человеческого уха лежит в пределах 20—20000 Гц, по теореме Котельникова частота семплирования должна быть как минимум в два раза больше максимальной частоты, которую может потребоваться передать — более 40 кГц.

- Кроме того, сигнал перед семплированием должен пройти через НЧ фильтр (иначе возникнет алиасинг) и, в то время как идеальный НЧ фильтр абсолютно не пропускал бы частоты выше 20 кГц, но полностью бы пропускал всё что ниже 20 кГц.
- На практике необходима т.н. переходная полоса, в которой происходит спад АЧХ (частотные составляющие подавляются лишь частично). Чем шире эта полоса, тем проще создать антиалиазинговый фильтр. Частота 44.1 кГц обеспечивает переходную полосу шириной 2.05 кГц.
- В ранние годы цифровой звук записывался на пленку видеокассет, т.к. это был единственный доступный носитель с ёмкостью достаточной для более-менее продолжительной записи звука. Чтобы свести к минимуму необходимые модификации оборудования, аудио воспроизводилось на той же скорости, что и видео; также использовалась практически идентичная схемотехника. Частота 44.1 кГц была признана наибольшей доступной.

Данная частота рассчитывается следующим образом:

- NTSC:

$$245 \times 60 \times 3 = 44100$$

245 активных строк/поле \times 60 полей/с \times 3 отсчета/строку = 44100
отсчетов/с

(490 активных строк на один кадр из 525 строк всего)

- PAL:

$$294 \times 50 \times 3 = 44100$$

294 активных строк/поле \times 50 полей/с \times 3 отсчета/строку = 44100
отсчетов/с

(588 активных строк на один кадр из 625 строк всего)

- Для существенного сокращения избыточности аудиоинформации применяются различные методы линейной и нелинейной обработки звуковых сигналов, которые приводят к сжатию с потерями, уменьшающими размер кодированной последовательности по сравнению с оригинальным за счет удаления информации, невоспринимаемой человеком; технология сжатия с потерями недостаток человеческого слуха превращает в преимущество, отбрасывая «ненужную» информацию; компромисс между малым цифровым потоком и качеством воспроизводимого аудиосигнала достигается путем изменения количества отбрасываемой информации
- Дополнительные проблемы в кодировании аудиоинформации возникают при обработке различных форматов – от стереофонического сигнала до объемного многоканального звукового сигнала; в этих случаях для существенного сокращения избыточности многоканальных сигналов используются корреляционные связи между ними

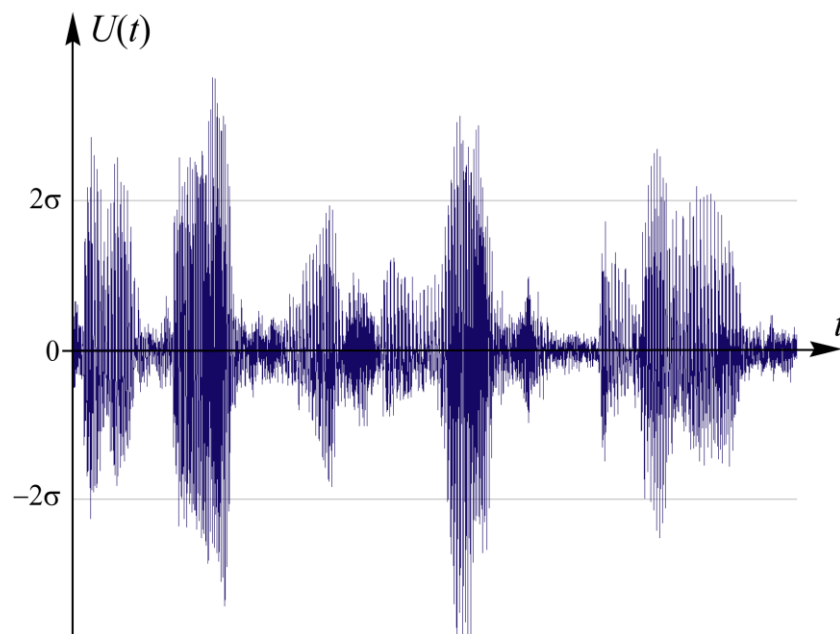
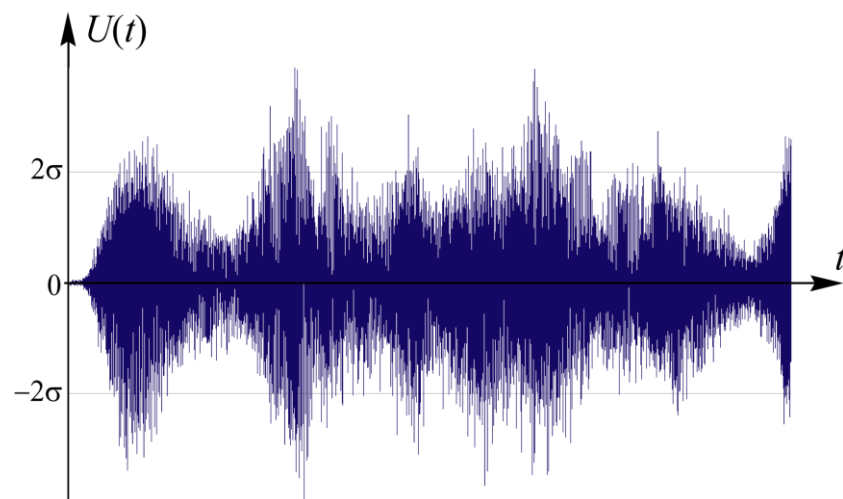
Основные характеристики звуковой информации

- Аналоговое представление звуковых сигналов основано на подобии форм и основных характеристик соответствующих им электрических сигналов
- В терминах теории информации, количество информации в электрическом сигнале в точности равно количеству информации в сигнале исходном, и электрическое представление не содержит избыточности, которая могла бы защитить переносимый сигнал от искажений при хранении, передаче и усилении
- Важные характеристики распространения звукового сигнала – интерференция, дифракция, рефракция, ревеберация, эхо, резонанс, диффузность, эффект Допплера и др.
- Звуковой сигнал или его электрический эквивалент $u(t)$ обычно считают случайным процессом с распределением мгновенных значений, которое характеризуется некоторой плотностью вероятностей $W(u)$

Основные характеристики звуковой информации

- **интерференция** – усиление колебаний звука в одних точках пространства и ослабление колебаний в других точках в результате наложения двух или нескольких звуковых волн
- **рефракция** – преломление, изменение направления движения звуковой волны от границы раздела с иной средой, поглощение или переход в другую среду
- **реверберация** – отражения звуковых колебаний в замкнутом пространстве, вызывающие специфический гул, изменяя тембральную окраску, насыщенность, глубину воспринимаемого звука
- **дифракция** – способность звуковых волн огибать препятствия
- **эхо** – возникновение сдвинутых во времени и различаемых отдельно повторов кратковременных звуковых колебаний
- **эффект резонанса** – способность звуковой волны, создаваемой некоторым колеблющимся телом, переносить энергию колебаний другому телу, которое поглощая эту энергию само становится источником звука
- **индекс диффузности** – фактор изотропности и однородности звукового поля
- **звуковое давление** – характеристика громкости звука, непосредственно воспринимаемого ухом человека
- **эффект Допплера** – изменение длины звуковых волн при изменении скорости движения слушателя относительно источника звука

Сигналограммы фрагментов музыкальной записи и речи



- Воспроизведение звучания больших симфонических оркестров, хоров, оркестров с хором соответствует нормальному распределению плотности вероятностей:

$$W(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right)$$

- Звучание отдельных музыкальных инструментов, легкой музыки соответствует распределению Лапласа:

$$W(u) = \frac{1}{\sigma\sqrt{2}} \exp\left(-\frac{|u|\sqrt{2}}{\sigma}\right)$$

- Звучание речи соответствует усредненному эмпирическому закону:

$$W(u) = \frac{1}{\sigma} \left[\exp\left(-\frac{3.31|u|}{\sigma}\right) + 0.18 \exp\left(-\frac{0.91|u|}{\sigma}\right) \right]$$

- σ - среднеквадратичное отклонение сигнала

Относительные плотности вероятностей мгновенных значений звуковых сигналов

(1) Воспроизведение звучания больших симфонических оркестров, хоров, оркестров с хором соответствует нормальному распределению плотности вероятностей:

$$W(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right)$$

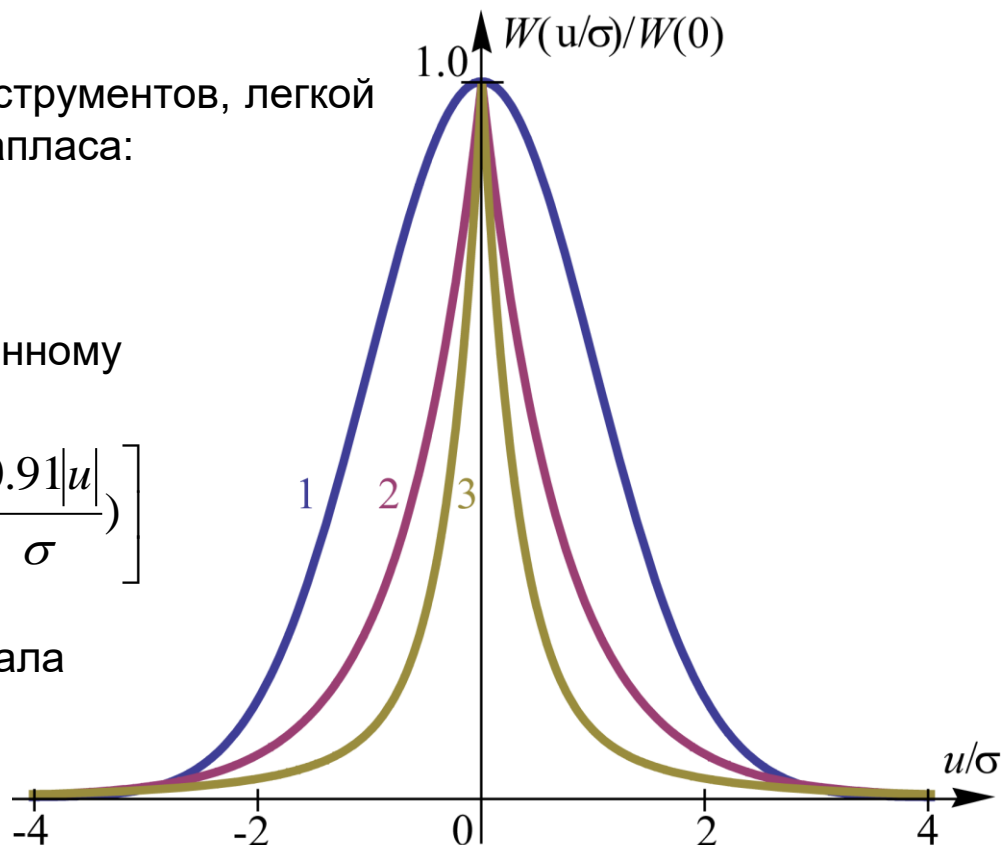
(2) Звучание отдельных музыкальных инструментов, легкой музыки соответствует распределению Лапласа:

$$W(u) = \frac{1}{\sigma\sqrt{2}} \exp\left(-\frac{|u|\sqrt{2}}{\sigma}\right)$$

(3) Звучание речи соответствует усредненному эмпирическому закону:

$$W(u) = \frac{1}{\sigma} \left[\exp\left(-\frac{3.31|u|}{\sigma}\right) + 0.18 \exp\left(-\frac{0.91|u|}{\sigma}\right) \right]$$

σ - среднеквадратичное отклонение сигнала



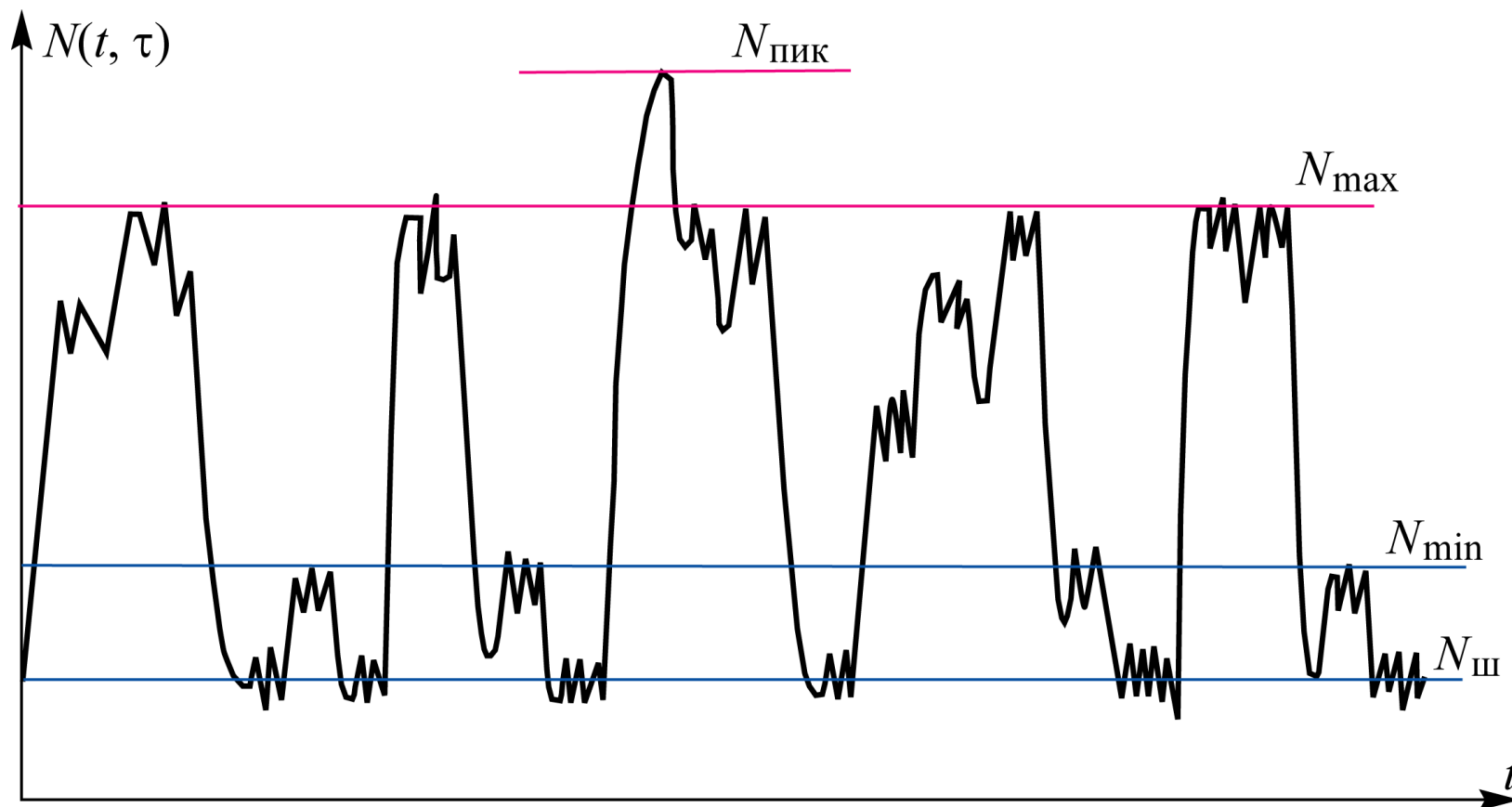
- Уровень электрического эквивалента звукового сигнала обычно характеризуют напряжением, формируемым на выходе квазипикового детектора с малой величиной постоянной времени заряда (порядка 5-10 мс) и значительной величиной τ постоянной времени разряда (1-2 с)
- Отношение усредненной величины выпрямленных мгновенных значений сигнала $U(t, \tau)$ на выходе квазипикового детектора (или выделяемой на нагрузке мощности $P(t, \tau)$ сигнала) к некоторой условной величине U_0 (или P_0) определяют формулой:

$$N(t, \tau) = 20 \lg \frac{U(t, \tau)}{U_0} = 10 \lg \frac{P(t, \tau)}{P_0}$$

- U_0 – принятая за начало отсчета среднеквадратическая величина электрического сигнала с эффективным напряжением 0,775 В на нагрузке 600 Ом (выделяемая мощность – $P_0=1$ мВт); выраженная в децибелах эта величина определяет значение уровня, равное 0 дБ
- Временная зависимость величины $N(t, \tau)$ – уровнеграмма сигнала $u(t)$ – характеризует изменение эффективного напряжения электрического эквивалента звукового сигнала
- Закон распределения уровней $W(N)$ при больших значениях постоянной времени τ практически мало зависит от характера звуковой программы и чаще всего соответствует нормальному распределению

Уровнеграмма звукового сигнала

- $N_{\max} - N_{\min}$ – диапазон квазимаксимального изменения уровня звукового сигнала
- $N_{\text{ш}}$ и $N_{\text{пик}}$ – уровень шума и случайный выброс квазипикового уровня в звуковом сигнале



Интенсивность звука или звуковое давление

- Интенсивность звука или звуковое давление оценивают либо в Паскалях, либо в децибелах относительно некоторого порога, величина которого принята равной $p_0 = 2 \cdot 10^{-5} \text{ Па} = 20 \text{ мкПа}$ и соответствует порогу слышимости здорового молодого человека в диапазоне звуковых частот 1-4 кГц
- Для характеристики уровней звукового давления (SPL – Sound pressure level) используется уравнение:

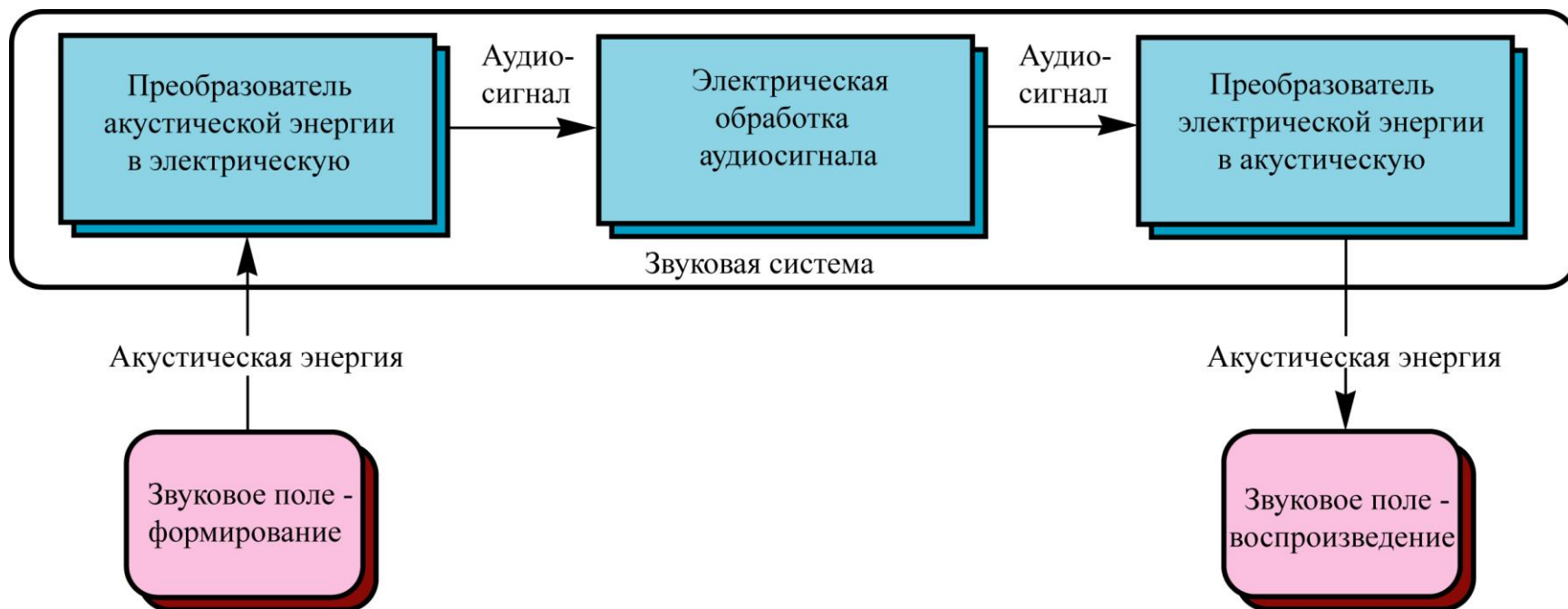
$$SPL = 20 \lg \frac{p_k}{p_0}$$

p_0 и p_k – звуковое давление, выраженное, например, в Паскалях

Уровни звукового давления, характерные для различных ИСТОЧНИКОВ 17

Источник звука	Звуковое давление, Па	Уровень звукового давления, дБ
Слуховой порог	$2 \cdot 10^{-5}$	0
Шелест листвы, слабого ветра, спокойное дыхание, тиканье наручных часов	$6 \cdot 10^{-5} - 6 \cdot 10^{-4}$	10 – 30
Шум в помещении, тихий разговор, скрипка пьанисимо, тихая музыка	$6 \cdot 10^{-4} - 6 \cdot 10^{-3}$	30 – 50
Разговор средней громкости, звук телевизора, шум в магазине, шум в ресторане	$6 \cdot 10^{-3} - 2 \cdot 10^{-2}$	40 – 60
Громкий разговор, легковой автомобиль, шумная улица, шум в машинописном бюро	$2 \cdot 10^{-2} - 2 \cdot 10^{-1}$	60 – 80
Грузовой автомобиль, движение на занятом шоссе, шум в метро во время движения, авторемонтная мастерская	$2 \cdot 10^{-1} - 6 \cdot 10^{-1}$	80 – 90
Женское и мужское пение	$6 \cdot 10^{-3} - 6 \cdot 10^{-1}$	40 – 90
Эстрадный и симфонический оркестры	$2 \cdot 10^{-2} - 6$	40 – 110
Громкая дискотека, пневматический молот, самолет на взлете, кузнечный цех	2 – 20	100 – 120
Реактивный самолет с расстояния	6 – 200	110 – 140
Порог болевого ощущения	60	130

- Принцип преобразования акустической энергии в электрическую, обработки электрического эквивалента и его последующего преобразования вновь в акустическую энергию поясняется на модели простейшей звуковой системы
- Источник звукового сигнала формирует акустическую энергию, которая преобразуется в колебания электрического аудиосигнала; после электрической обработки сигнала, включающей, например, усиление, формирование спецэффектов, изменение динамического диапазона и др., аудиосигнал подвергается обратному преобразованию в акустическую энергию и ее воспроизведению



Уровни электрического эквивалента звуковой системы (N , дБ) пропорциональны уровням звукового давления (SPL , дБ)

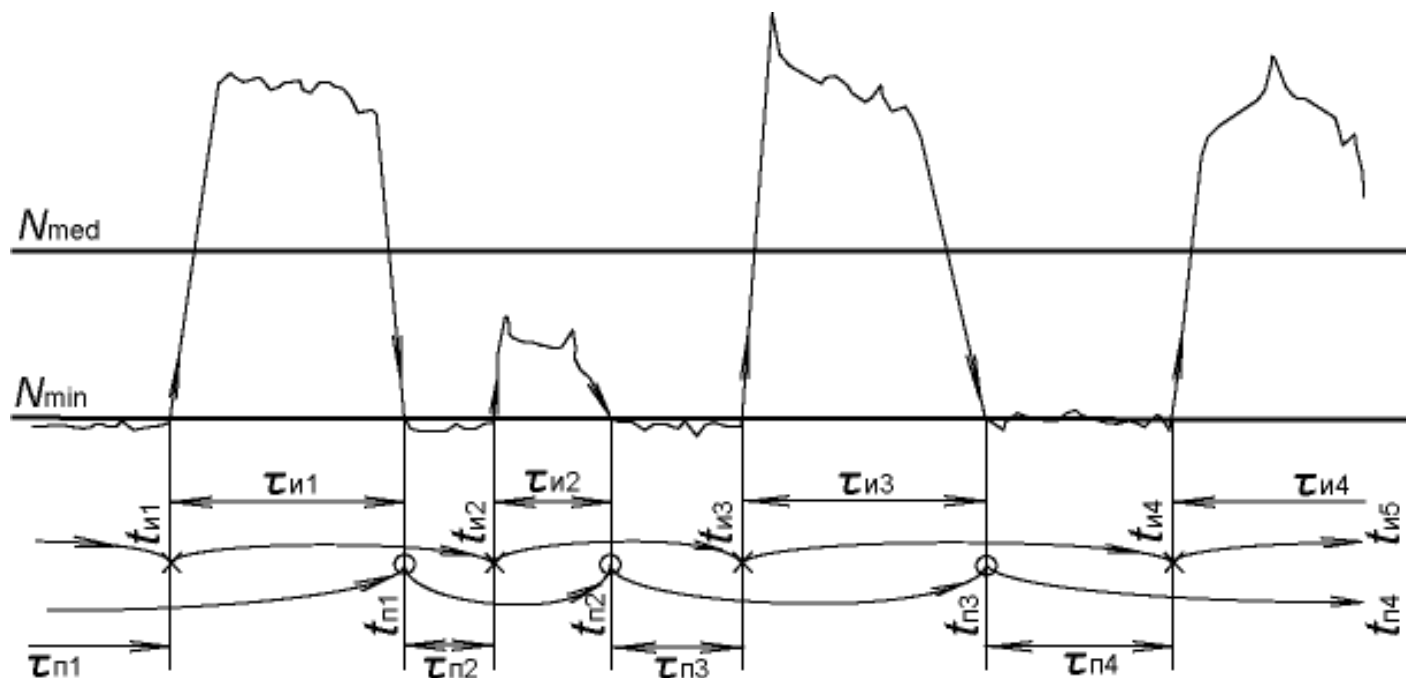
При уровне $SPL=120$ дБ звукового давления на входе преобразователя звука в электрический эквивалент максимальный уровень электрического сигнала может достигать величины $N=+25$ дБ

Соответствие этих величин представлено на рисунке (между величинами SPL и N имеется линейная зависимость, если в системе не применяется компрессия, эквалаизация, ограничение или отсечка сигналов) Такое соотношение характерно для вещательных систем любого типа, звукоусиления, звукозаписи



- Требования к динамическому диапазону звуковой системы зависят от ее назначения и области использования
- Динамический диапазон системы в звуконепроницаемой студии звукозаписи может быть большим, поскольку в таком случае шумы в не превышают 10 – 15 дБ
- Звуковая система, предназначенная для усиления симфонической музыки, должна иметь запас динамического диапазона более 20 дБ, так как пиковые значения, соответствующие звукам некоторых инструментов (литавр, скрипок и др.) могут достигать 120 дБ
- В системах, предназначенных только для воспроизведения речи или предупреждающих сигналов, уровень звука можно контролировать и удерживать в очень узком диапазоне

- Речевой сигнал можно рассматривать как последовательность импульсов, разделенных паузами, при которых уровень сигнала ниже некоторого минимального уровня N_{\min}
- $T_{и1}, T_{и2}, \dots$ – интервалы последовательности импульсов
 $T_{п1}, T_{п2}, \dots$ – интервалы последовательности пауз
 $t_{и1}, t_{и2}, \dots$ – времена переходов от пауз к импульсам
 $t_{п1}, t_{п2}, \dots$ – времена переходов от импульсов к паузам



Спектральные характеристики звуковых сигналов

- Реальные звуковые сигналы практически невозможно описать какой-либо математической функцией или эмпирической зависимостью. По этой причине как правило анализируется лишь ограниченные во времени фрагменты звукового сигнала, ограниченные некоторой оконной функцией. При этом используется понятие «мгновенный спектр»:

$$F(f, \tau) = \int_{-\tau/2}^{+\tau/2} w(t) \cdot u(t) \cdot \exp(-j2\pi ft) dt$$

- где $u(t)$ – электрический эквивалент звукового сигнала, $w(t)$ – оконная функция, равная нулю вне пределов заданного интервала $-\tau/2 \leq t \leq \tau/2$, f – текущая частота
- Часто оценивают спектральную плотность мощности аудиосигнала с применением относительно узкополосного фильтра с полосой пропускания Δf :

$$G(f_0, \tau) = \frac{2}{\Delta f \cdot \tau} \int_{f_0 - \Delta f/2}^{f_0 + \Delta f/2} |F(f, \tau)|^2 df$$

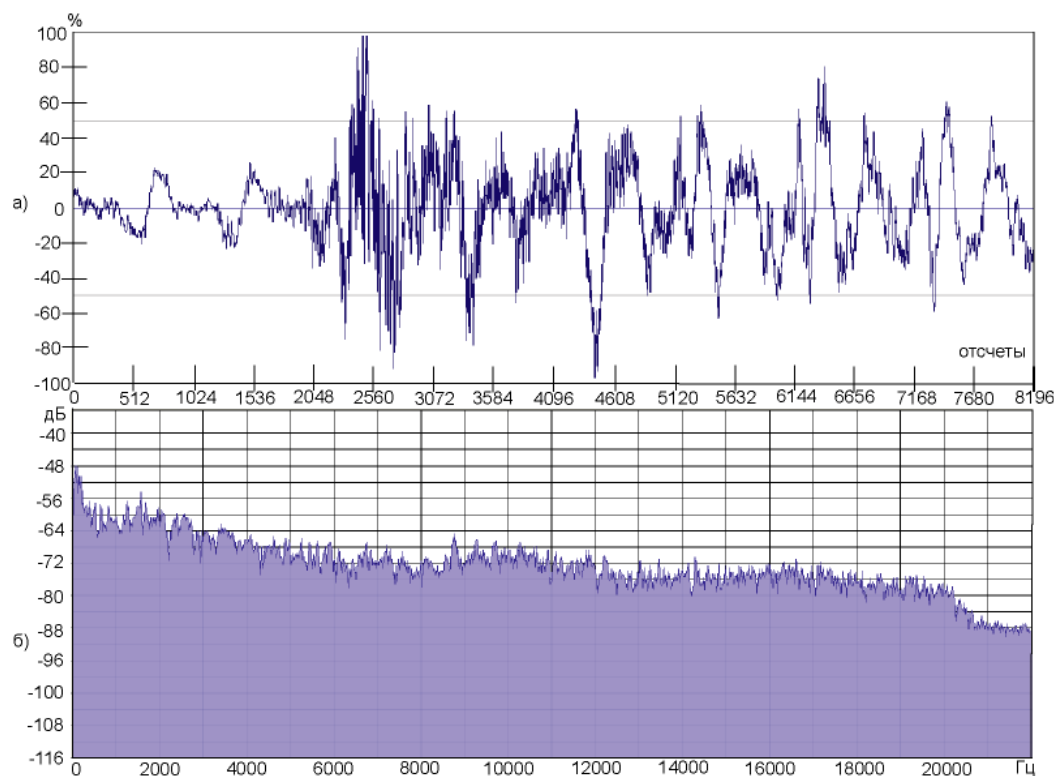
- Результат спектрального анализа выражают в децибелах (G_0 – значение, соответствующее нулю шкалы уровней):

$$N(f_0, \tau) = 10 \lg \frac{G(f_0, \tau)}{G_0}$$

- Для расчета частотного спектра ограниченного по длительности сигнала и представленного его дискретными значениями используют дискретное преобразование Фурье (ДПФ)

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} = \sum_{n=0}^{N-1} x_n \cdot [\cos(2\pi kn/N) - i \cdot \sin(2\pi kn/N)], \quad (k = 0, \dots, N-1).$$

- или его разновидность – быстрое преобразование Фурье (БПФ)



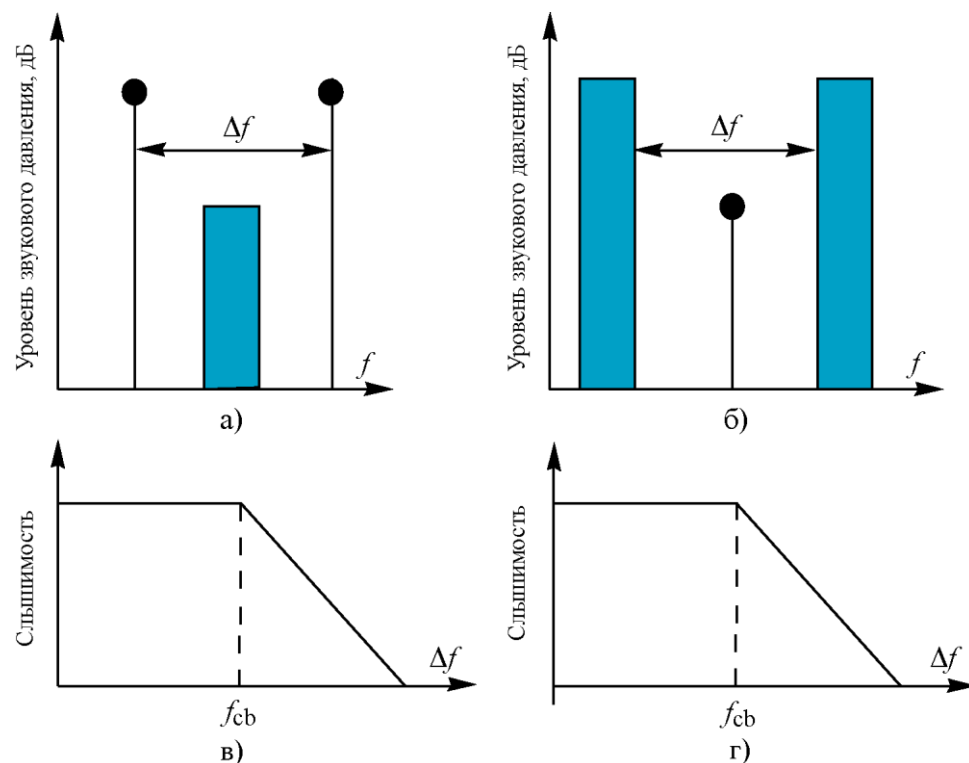
- С точки зрения обработки электрического эквивалента звука и его передачи весьма полезной является информация о спектрах разговорных и певческих голосов, музыкальных инструментов, природных и индустриальных шумов и т.д. Сведения о частотных диапазонах некоторых источников звука даны в табл.

Источник звука	Граничная частота, Гц	
	нижняя	Верхняя
Мужской голос	100	7000
Женский голос	200	9000
Рояль	100	5000
Скрипка	200	14000
Флейта	250	14000
Тарелки	400	12000
Литавры	65	3000
Шум шагов	100	10000
Аплодисменты	150	15000

- Специфической особенностью восприятия звука является разделение его спектра на полосы равной разборчивости (критические полосы), когда сигнал в пределах одной и той же полосы создает весьма близкие слуховые ощущения (различные участки в улитке внутреннего уха «настроены» на разные частотные полосы)
- Экспериментально ширина критической полосы может быть определена по резкому уменьшению субъективной громкости (на рис. схематично показаны методы такого определения)

• Методы измерения критической полосы:

- (а, в) узкополосный шумовой компонент маскируется двумя тональными компонентами, порог обнаружения шумовой компоненты резко уменьшается при разности частот тональных компонент за пределами критической полосы
- (б, г) тональные и шумовые компоненты меняются ролями



- В частотном промежутке от 0 до 16 кГц опытным путем определены 24 полосы, оцениваемые в Барках (по фамилии немецкого ученого Баркгаузена Г.Г., 1881-1956)

Номер полосы, Барк	Диапазон критической полосы, Гц	Ширина критической полосы, Гц	Центральная частота критической полосы, Гц
0	0 – 100	100	50
1	100 – 200	100	150
2	200 – 300	100	250
3	300 – 400	100	350
4	400 – 510	110	450
5	510 – 630	120	570
6	630 – 770	140	700
7	770 – 920	150	840
8	920 – 1080	160	1000
9	1080 – 1270	190	1170
10	1270 – 1480	210	1370
11	1480 – 1720	240	1600
12	1720 – 2000	280	1850
13	2000 – 2310	320	2150
14	2310 – 2700	380	2500
15	2700 – 3150	450	2900
16	3150 – 3700	550	3400
17	3700 – 4400	700	4000
18	4400 – 5300	900	4800
19	5300 – 6400	1100	5800
20	6400 – 7700	1300	7000
21	7700 – 9500	1800	8500
22	9500 – 12000	2500	10500
23	12000 - 15500	3500	13500

- Ширина критической полосы остается примерно постоянной (около 100 Гц) вплоть до значения центральной частоты полосы 500 Гц, а при больших значениях увеличивается в среднем на 20% центральной частоты
- В работах по психоакустике используется следующая аппроксимация этой зависимости:

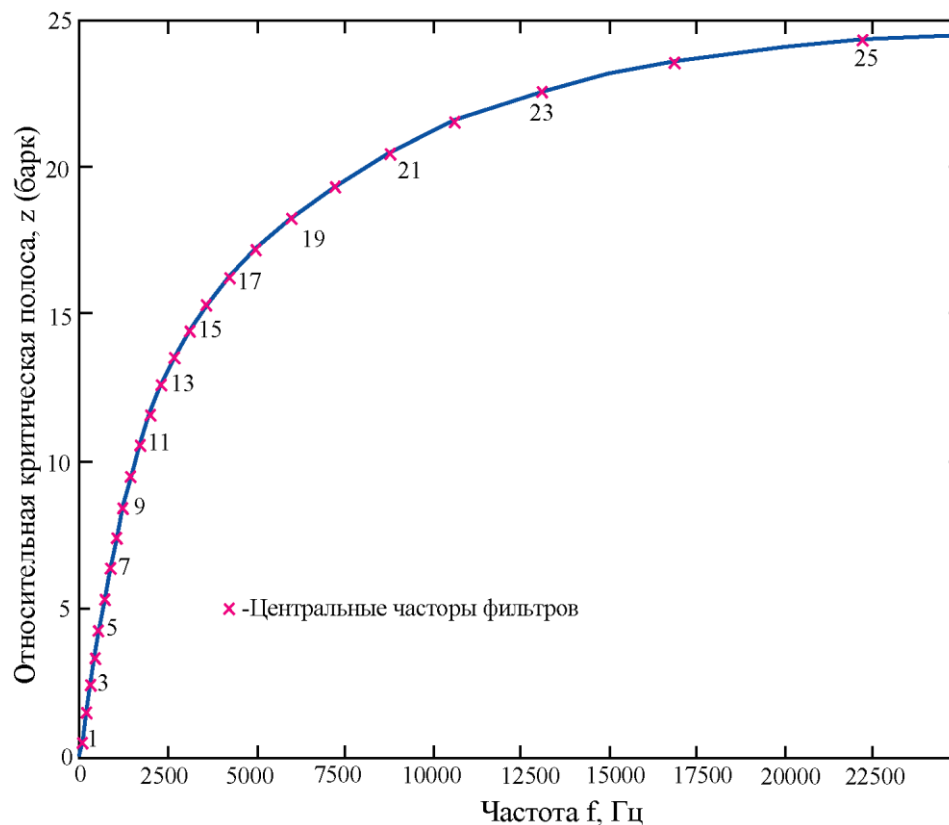
$$BW(f) = 25 + 75 \left[1 + 1,4(f / 1000)^2 \right]^{0,69}, \quad \text{Гц}$$

- Хотя функция непрерывна, для практических задач лучше использовать дискретный набор полосовых фильтров, покрывающий всю рабочую полосу частот

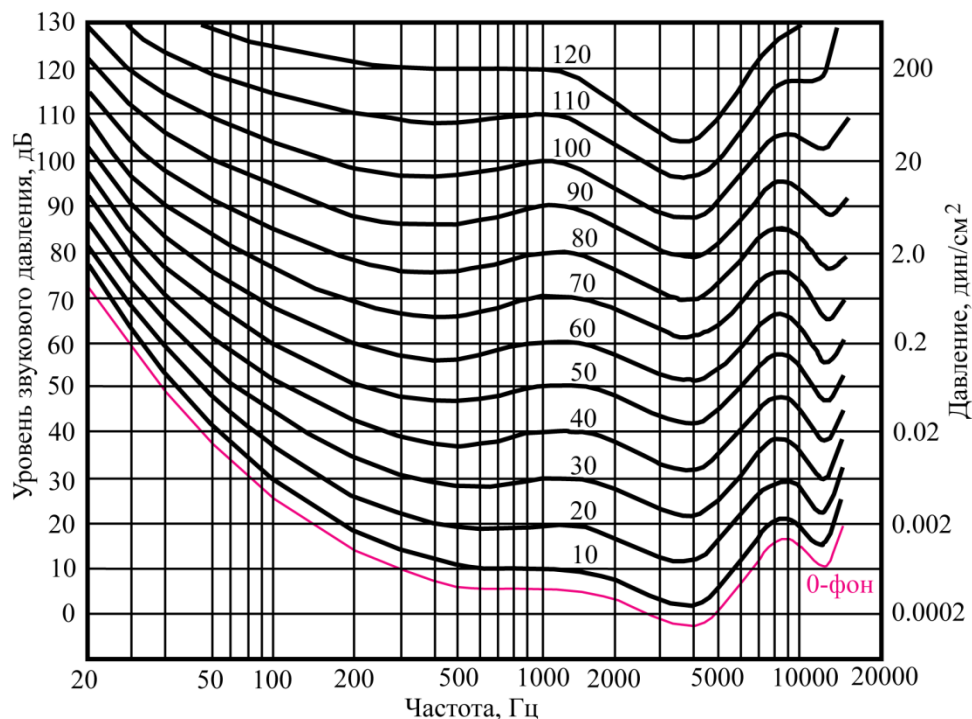
№ полосы	Центр. частота (Гц)	Ширина полосы (Гц)	№ полосы	Центр. частота (Гц)	Ширина полосы (Гц)	№ полосы	Центр. частота (Гц)	Ширина полосы (Гц)
1	50	-100	10	1175	1080-1270	19	4800	4400-5300
2	150	100-200	11	1370	1270-1480	20	5800	5300-6400
3	250	200-300	12	1600	1480-1720	21	7000	6400-7700
4	350	300-400	13	1850	1720-2000	22	8500	7700-9500
5	450	400-510	14	2150	2000-2320	23	10500	9500-12000
6	570	510-630	15	2500	2320-2700	24	13500	12000-15500
7	700	630-770	16	2900	2700-3150	25	19500	15500-
8	840	770-920	17	3400	3150-3700			
9	1000	920-1080	18	4000	3700-4400			

- Для определения расстояния между центральными частотами соседних критических полос используют специальную единицу частоты: 1 барк
- На рис. показана связь между частотой, выраженной в герцах и барках
- Эта зависимость используется для преобразования линейной шкалы частот в шкалу, соответствующую спектральному восприятию человека, для ее аппроксимации было применено следующее выражение (f – частота в [Гц]):

$$z(f) = 13 \arctg(0,00076f) + 3.5 \arctg \left[\left(\frac{f}{7500} \right)^2 \right], \text{ Барк}$$

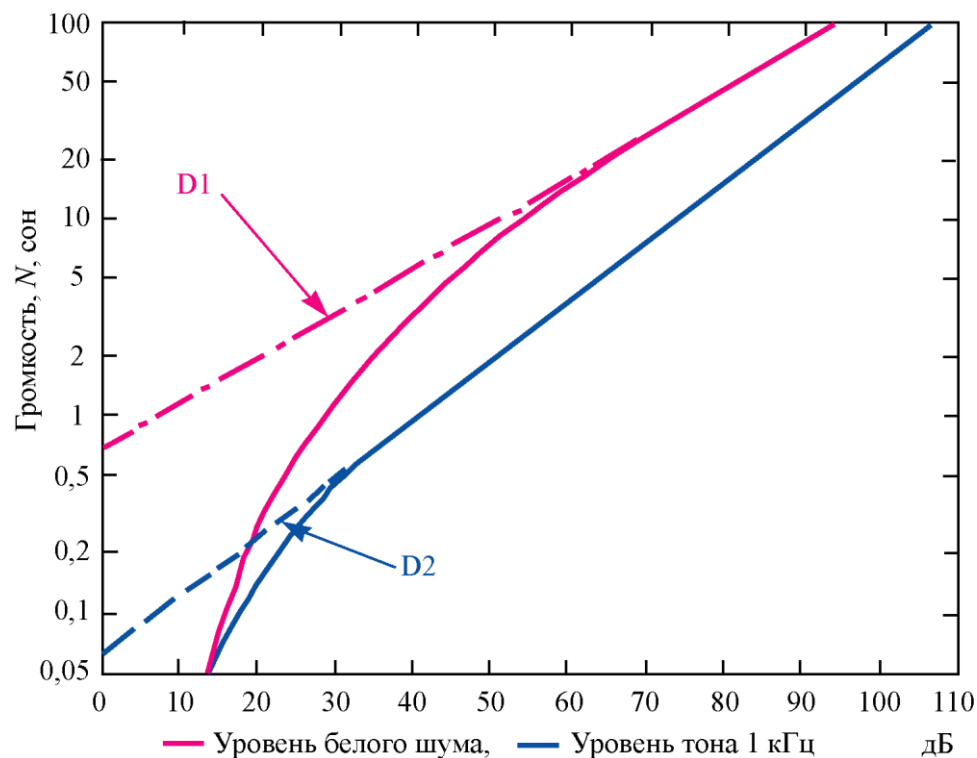


- Одной из наиболее важных характеристик восприятия звука является громкость, которая характеризует интенсивность звукового события
- Громкость есть характеристика субъективного восприятия, на которую помимо амплитуды звуковой волны оказывают влияние различные другие факторы
- Совершенно по-разному воспринимается громкость в зависимости от частоты звука
- Этот факт был впервые установлен двумя исследователями, Флетчером и Мэнсоном, которые разработали набор графиков, получивших название кривых равной громкости, или кривых Флетчера-Мэнсона
- Каждая кривая на этом графике характеризует уровень равной громкости с начальной точкой отсчета на частоте 1000 Гц



- Кривые равной громкости (изофоны) были получены учеными Флетчером и Мэнсоном в результате обработки данных большого числа экспериментов, проведенных ими среди нескольких сотен посетителей Всемирной выставки 1931 года в Нью-Йорке
- В настоящее время в международном стандарте ISO 226 (1987 г.) приняты уточненные данные измерений, полученные в 1956 году
- Изофоны представляют собой графики зависимостей уровня звукового давления от частоты при заданном уровне громкости
- Фон - единица для оценки уровня громкости звука
- С помощью этих кривых можно определить уровень громкости чистого тона какой-либо частоты, зная уровень создаваемого им звукового давления
- Самая нижняя кривая примерно соответствует громкости 3 фон и описывает абсолютный порог слышимости ($ATH [1000 \text{ Гц}] = 3,369 \text{ дБ SPL}$)
- Кривые равной громкости показывают, какая требуется разница в уровне звукового давления, чтобы звуки всех частот воспринимались с такой же громкостью, как референсный синусоидальный сигнал с частотой 1 кГц
- Числа, находящиеся над каждой кривой, представляют собой меру громкости, выраженную в фонах, на референсной частоте 1 кГц фоны равны децибелам, т.е. у чистого тона с частотой 1 кГц уровень в фонах численно равен уровню звукового давления в децибелах
- К примеру, при очень низком уровне громкости 30 фон (30 дБ SPL на частоте 1 кГц) басовый тон 50 Гц должен воспроизводиться с уровнем 60 дБ SPL, чтобы он воспринимался с такой же громкостью, как звук 30 дБ SPL с частотой 1 кГц, т.е. ухо человека менее чувствительно в области низких и высоких частот, нежели на средних частотах

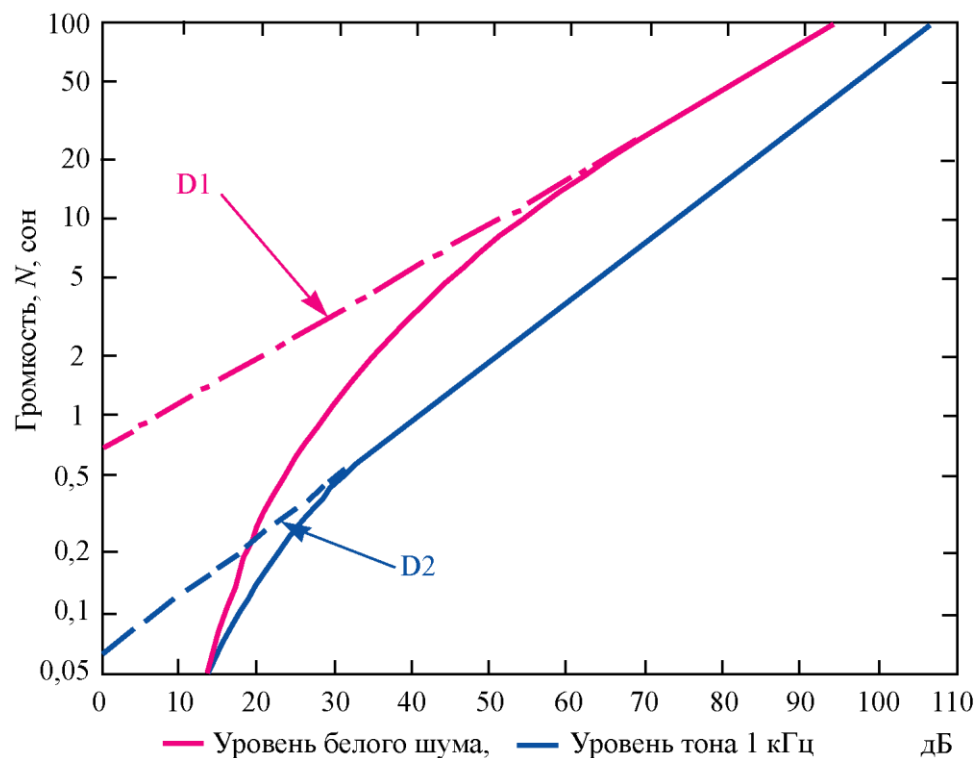
- Уровень громкости может измеряться также в сонах
- Преимущество оценки уровней в сонах состоит в том, что для звука из нескольких компонентов, сильно разнесенных по частоте, общий уровень громкости в сонах, равен сумме уровней каждой из компонентов
- Существует однозначная связь между уровнями громкости чисто тонального звука в сонах и фонах
- На рис. показана зависимость уровня громкости в сонах тонального звука частоты 1 кГц (сплошная линия) и белого шума (точечная линия) от уровня звукового давления
- Уровень 1 сон соответствует громкости чистого тона частоты 1000 Гц с уровнем 40 дБ
- При увеличении уровня громкости на 10 фонов громкость звука в сонах возрастает в 2 раза, это значит, что уровням громкости 40, 50 и 60 фон соответствуют громкости 1, 2 и 4 сон
- На уровнях ниже 40 фон, изменение уровня в сонах происходит намного быстрее, уровень 3 фон, соответствующий абсолютному порогу слышимости, равен 0 сон



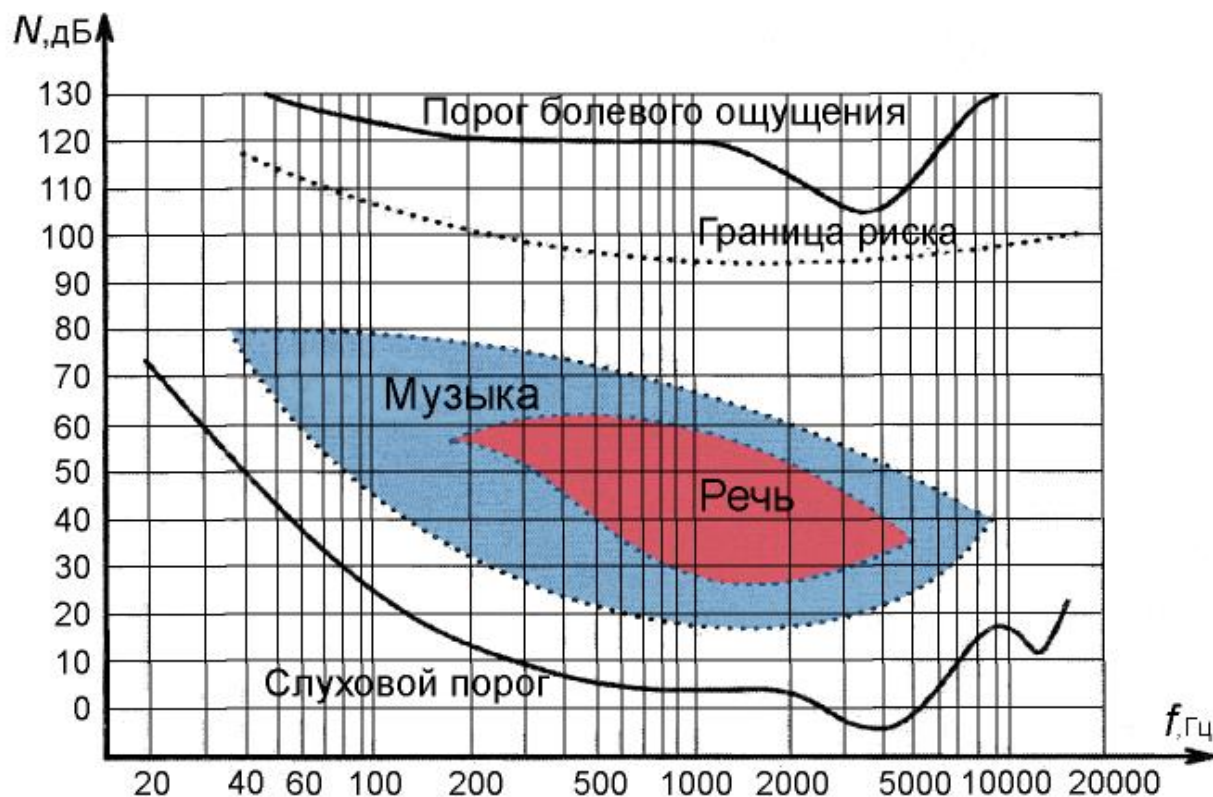
- При проведении экспериментов по субъективной оценке уровня громкости для белого шума устанавливался уровень звукового давления эталонного звука частоты 1000 Гц, затем испытуемому предлагался прослушать белый шум и отрегулировать его уровень таким образом, чтобы он казался равно громким эталонному
- Если условиться, что после регулировки белого шума и эталонного тонального звука уровни громкости в сонах совпадают, то получается зависимость уровня громкости белого шума в сонах от уровня звукового давления, которая изображена красной линией на рис.
- На этом рисунке также видно, что при большой громкости уровень эталонного звука в сонах пропорционален интенсивности в степени 0,3, а уровень громкости белого шума – от степени 0,23

$$D1[\text{сон}] = \frac{2}{3} \left(\frac{I_{\text{БШ}}}{I_0} \right)^{0,23}$$

$$D2[\text{сон}] = \frac{1}{16} \left(\frac{I_{1\text{кГц}}}{I_0} \right)^{0,3} = 2^{-4 + \frac{1}{10} L_{1\text{кГц}} [\text{дБ}]}$$



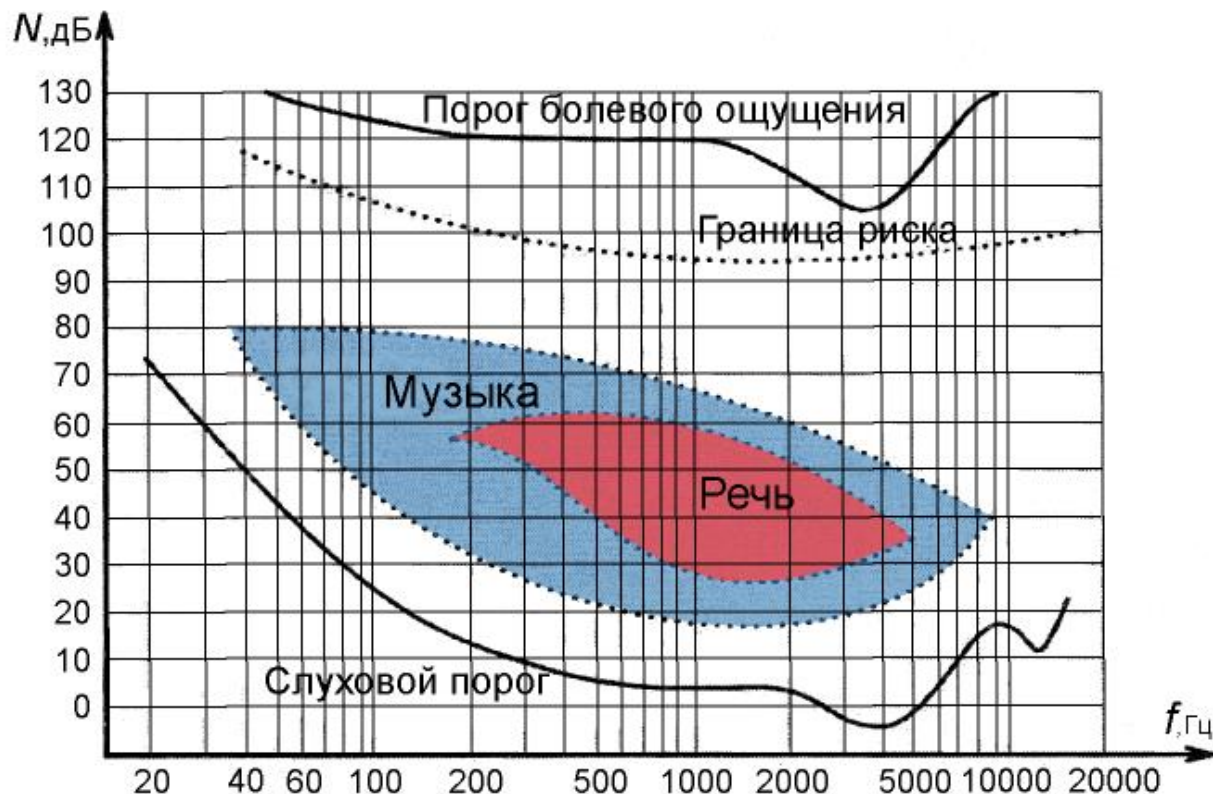
- Под абсолютным порогом слышимости (Absolute Threshold of Hearing, Threshold in Quiet) понимают минимальную энергию чистого тона, которая еще позволяет слушателю определить наличие этого тона при отсутствии окружающего шума
- Зависимость этой энергии от частоты приведена на рис., на котором также изображены порог болевого ощущения, граница риска, а также области передачи музыки и речи
- Области громкости и спектральных областей речи и музыки различны
- Приведенные графики определены для случая полной тишины, в случае иных условий, например, в шумной комнате, диаграммы окажутся другими



- Существование абсолютного порога слышимости можно объяснить наличием внутреннего шума, порождаемого в ухе человека
- Внутренний шум является отдельным маскером, который вызывает непрерывный порог маскирования воспринимаемого звука
- Для аппроксимации частотной зависимости абсолютного порога слышимости хорошо подходит следующая нелинейная функция:

$$ATH(f) = T_q(f) = 3,64 f^{-0,8} - 6,5 e^{-0,6(f-3,3)^2} + 10^{-3} f^4, \quad \text{dB SPL}$$

- f – частота в кГц, SPL - величина звукового давления (Sound Pressure Level)



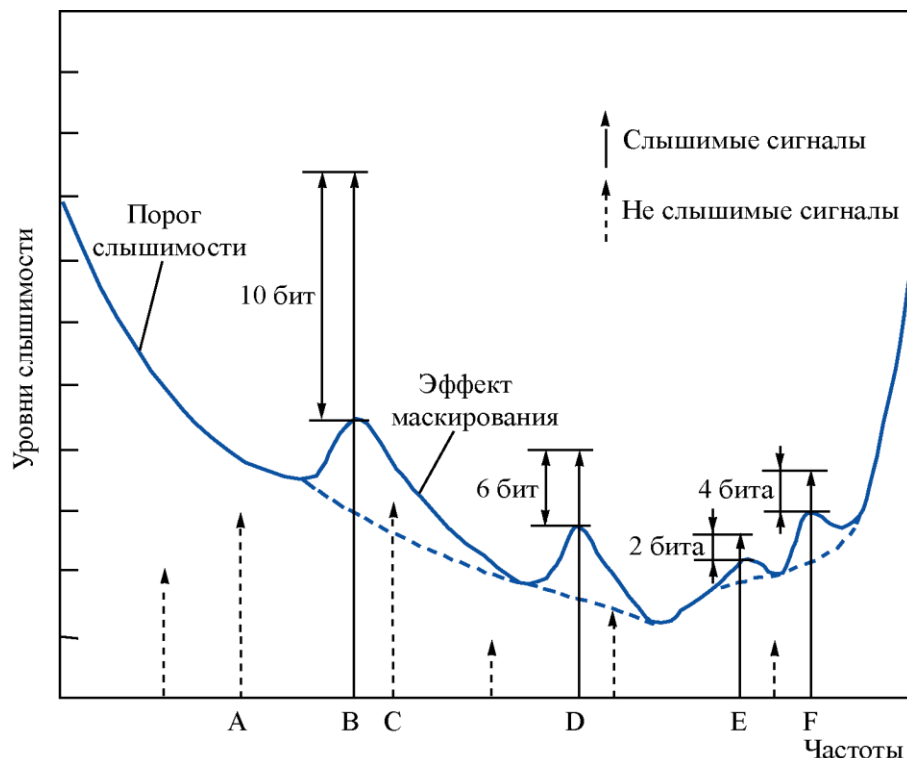
Разновидности шумов и их спектры

В профессиональной литературе рассматриваются несколько различных по спектру разновидностей шумов

- **Белый** шум обладает постоянной спектральной плотностью на всей протяженности спектра
- **Розовый** шум, спектральная плотность которого уменьшается на 3 дБ с каждой последующей октавой
- **Оранжевый** шум, спектральная плотность которого квазипостоянна и имеет полосы нулевой энергии, рассеянные на всей протяженности спектра; такие полосы располагаются около частот, соответствующих музыкальным нотам
- **Зеленый** шум подобен розовому шуму с усиленной областью в районе 500 Гц
- **Синий** шум, спектральная плотность которого увеличивается на 3 дБ с каждой последующей октавой
- **Фиолетовый** шум или дифференцированный белый шум, спектральная плотность которого увеличивается на 6 дБ с каждой последующей октавой
- **Серый** шум, спектр имеет форму, аналогичную графику психоакустической кривой порога слышимости
- **Коричневый** шум, спектральная плотность которого уменьшается на 6 дБ с каждой последующей октавой
- **Тональный** шум, в спектре которого имеются слышимые дискретные тоны
- **Черный** шум имеет постоянную конечную спектральную плотность за пределами частотного порога слышимости, равного 20 кГц

Частотное (одновременное) маскирование

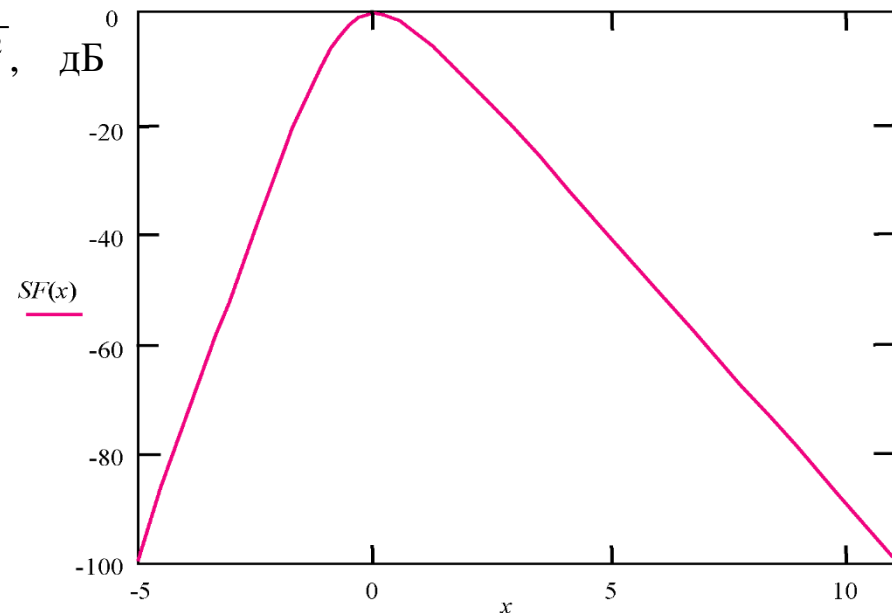
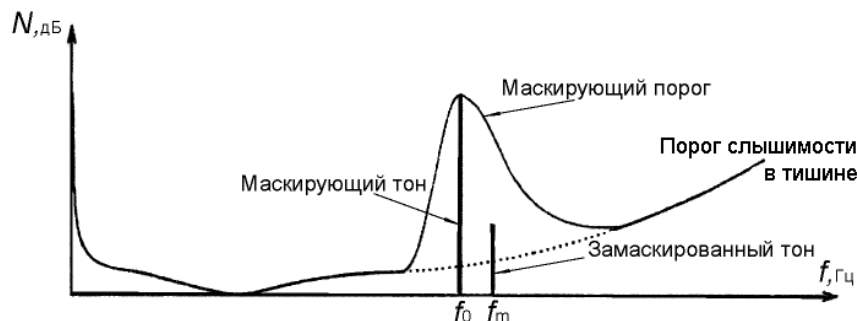
- С механизмом критических полос слуха человека связаны свойства межполосового и внутри полосового частотного маскирования
- Под маскированием понимают ситуацию, при которой один звук становится неслышимым из-за присутствия другого звука (см. рис.)
- С целью оптимального расчета порога маскирования следует различать два вида частотного маскирования: тон-шум и шум-тон
- В первом случае тональный сигнал, расположенный в центре критической полосы, маскирует шум в пределах ширины полосы или некоторой ее окрестности
- Во втором случае, наоборот, маскирующим сигналом является шум, а маскируемым — ТОН



Частотное (одновременное) маскирование

- Эффект маскирования упрощенно можно объяснить тем, что сильный тональный или шумовой masker создает очаг возбуждения на участке базилярной мембраны, соответствующем критической полосе, это возбуждение препятствует ощущению более слабого сигнала
- Порог маскирования снижается при увеличении разницы частот маскирующего и маскируемого сигналов
- Данное явление называют распространением маскирования, в алгоритмах кодирования часто моделируется треугольной функцией распространения (Spreading Function, SF) с наклоном +25 и -10 дБ/барк
- Для более точного представления функции распространения используется выражение (x – разность частот маскируемого и маскирующего сигналов в [барк]):

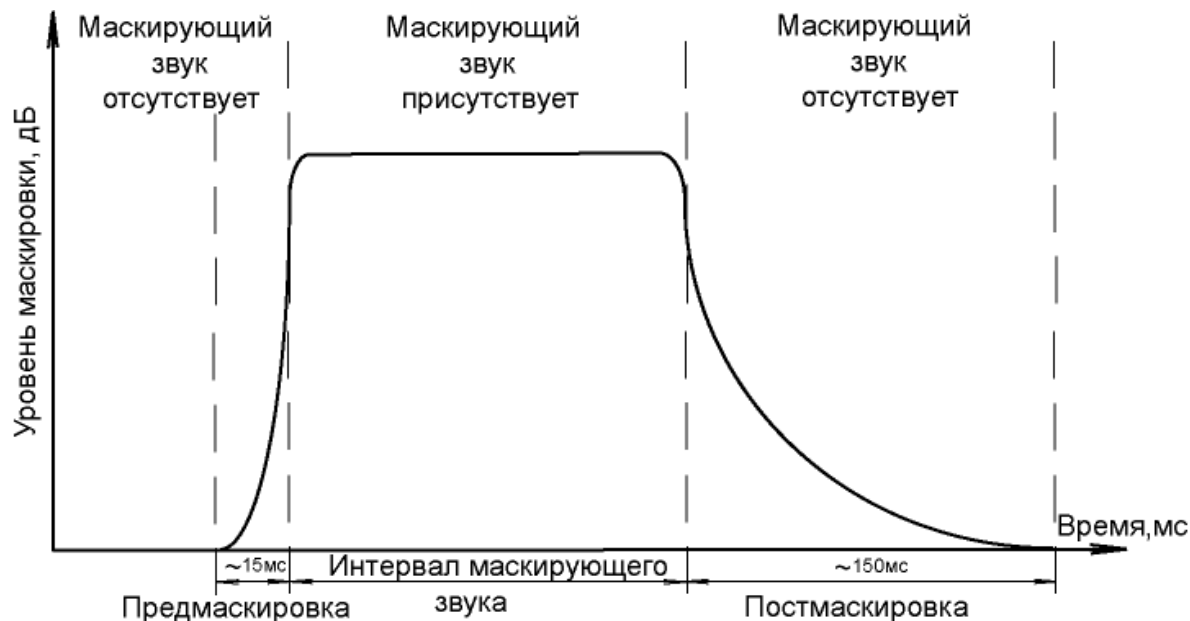
$$SF(x) = 15,81 + 7,5(x + 0,474) - 17,5\sqrt{1 + (x + 0,474)^2}, \text{ дБ}$$



Временное (неодновременное) маскирование

- Эффект частотной маскировки справедлив для частотных составляющих, присутствующих в спектре сигнала в одно и то же время
- За счет инерционности слуха эффект маскировки распространяется и во временной области
- В случае, когда маскирующий тон прекращается раньше маскируемого тона, создается постмаскируемый эффект
- Возможна и предмаскировка, когда маскирующий тон появляется несколько позже маскируемого тона
- Интервал предмаскировки (~ 5 мс) существенно меньше интервала постмаскировки ($\sim 50 - 300$ мс)

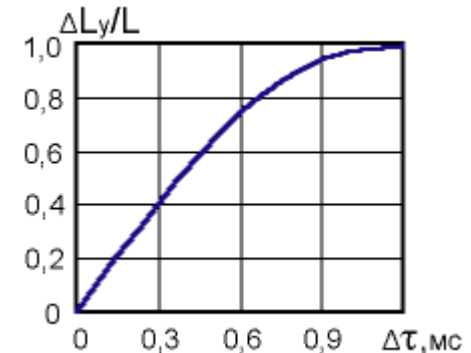
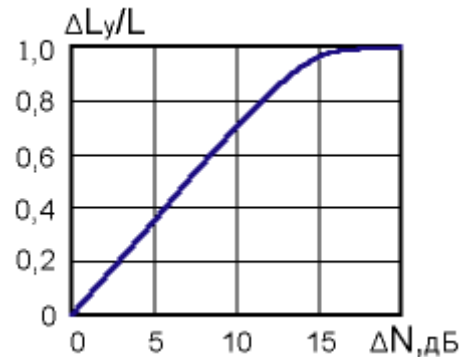
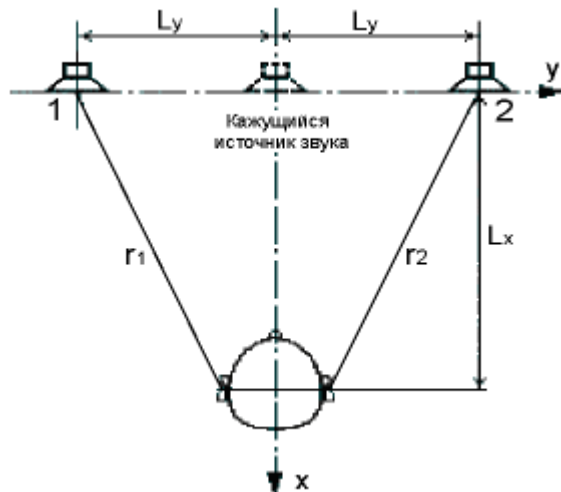
Эффекты маскировки, изменения порогов слышимости широко используются в самых различных технологиях цифрового сжатия аудиоинформации



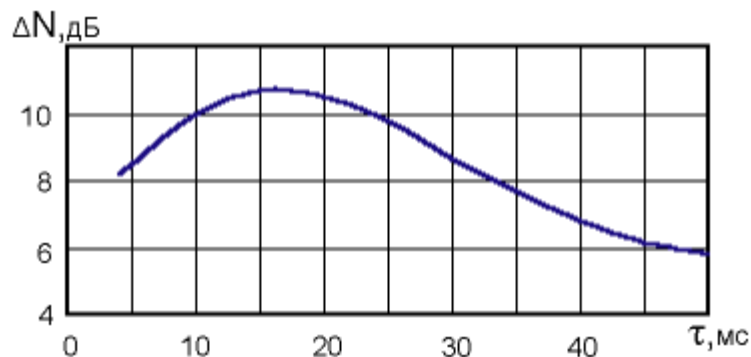
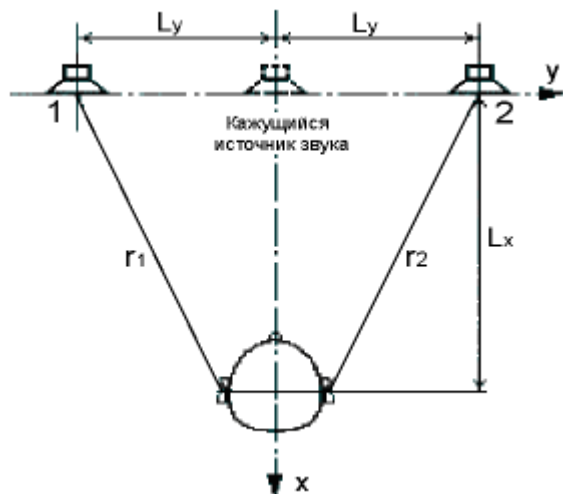
Пространственное восприятие звуковых сигналов

- Локационные способности восприятия звука, так называемый **бинауральный эффект**, объясняются фазовым смещением звуковых волн, неодинаковым уровнем звуковых давлений в ушах, особенностями тембров знакомых источников звуков и их изменений
- При неподвижном источнике звука слух способен определить направление движения звуковых волн по горизонтали не точнее 12 градусов, а по вертикали - 17...20 градусов
- Бинауральный эффект практически отсутствует на частотах ниже 300 Гц, на частотах от 300 до 1000 Гц становится заметным сдвиг фаз звуковых волн, попадающих в правое и левое ухо, при частотах более 1000 Гц сдвиг фаз становится очень небольшим и поиск направления осуществляется за счет сравнения силы звука, приходящего с разных сторон

- На рис. поясняется эффект интегральной локализации восприятия информации от двух источников звука
- Два одинаковых источника (1 и 2), расположенные на расстоянии $2L_y$ один от другого, а на расстоянии L_x расположен слушатель, уши которого находятся на расстоянии r_1 и r_2 от соответствующих источников
- Если на оба источника излучают звуковую энергию одинаковой мощности, то звук от каждого источника достигнет ушей одновременно и идентичность звуков не позволит слуху разделить их в пространстве, например, на левый и правый; при этом возникает иллюзия: виртуальный (кажущийся) источник звука как бы находится в середине между источниками звука
- Если уменьшить мощность излучения одного из источников, то это воспринимается как перемещение кажущегося источника в сторону второго источника; таким образом, варьируя громкость звучания левого и правого источника, можно вызывать и поддерживать иллюзию перемещения виртуального источника звука
- Аналогичная иллюзия перемещения виртуального источника возникает, если создать запаздывание звука в одного из источников
- Оба эти эффекта широко используются при цифровой обработке и записи музыки



- При задержках одного из сигналов на время более 50 мс наличие запаздывающего сигнала ощущается как помеха в виде эха. Опережающий сигнал при одинаковом уровне с задержанным в этом случае подавляет (маскирует) последний. Повышая уровень запаздывающего сигнала, можно добиться того, что оба источника звука будут восприниматься отдельно даже при запаздывании менее 50 мс
- На рисунке показано необходимое превышение уровня (ΔN , дБ) запаздывающего сигнала в зависимости от временной задержки. При $t = 15...20$ мс уровень задержанного сигнала должен быть повышен на 11 дБ, чтобы оба источника звука воспринимались отдельно. При $t < 50$ мс для этого эффекта достаточно превышение уровня всего на 6 дБ. При $t < 5$ мс наблюдается неустойчивый режим: виртуальный источник звука как бы перепрыгивает из одного источника в другой, совпадая то с источником опережающего, то с источником задержанного сигнала



- Для качественного восприятия реального пространственного звучания музыкальных программ использование двухканальной (стереофонической) системы воспроизведения звуковых сигналов не всегда является достаточным
- Основная причина этого кроется в том, что стерео сигнал, приходящий к слушателю от двух физических источников звука, определяет расположение мнимых источников лишь в той плоскости, в которой расположены реальные физические источники звука
- Поэтому в последние десятилетия стали развиваться системы многоканального воспроизведения звука, реализующие так называемый **трансуральный эффект**
- Для воссоздания более или менее реалистичного, действительно объемного звучания прибегают к применению сложных приемов, моделирующих особенности слуховой системы человека, а также физические особенности и эффекты передачи звуковых сигналов в пространстве. Главная проблема заключается в том, чтобы создать такой сигнал, который бы при помощи двух или более источников звука воспринимался слушателем как трехмерный

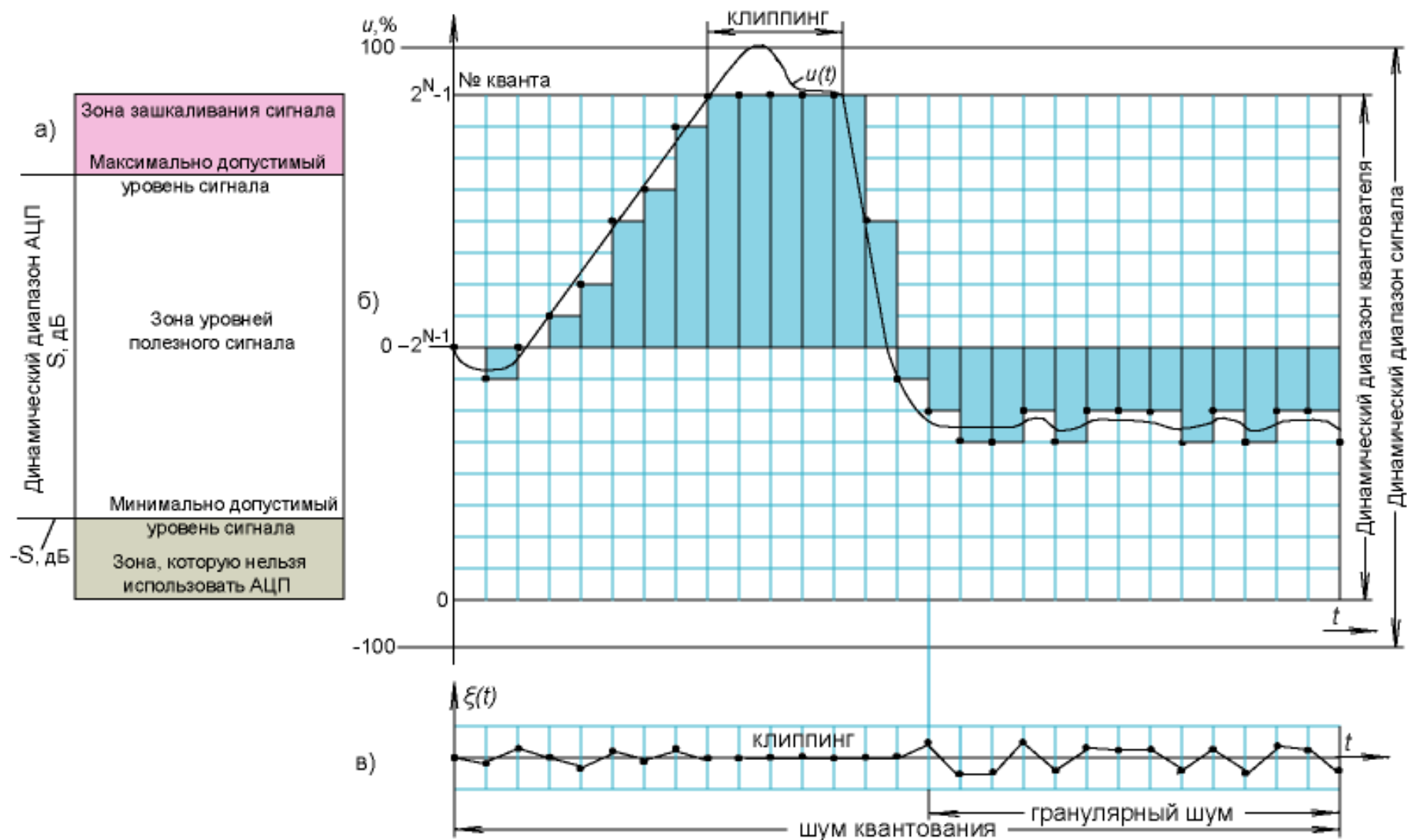
Основные принципы цифровых преобразований звуковых сигналов

- Преобразование аналогового звукового сигнала путем временной дискретизации и квантования выбранных дискретных его значений неизбежно приводит к невозможности его абсолютно точного восстановления
- Если предположить, что аналоговый сигнал строго ограничен по спектру, то в соответствии с теоремой Котельникова он полностью определяется дискретной последовательностью своих мгновенных значений, взятых с частотой выборки f_d как минимум вдвое превышающей граничную частоту спектра: $f_d \geq 2f_{гр}$
- Квантование дискретных отсчетов сигнала вызывает шумовые искажения, величина которых зависит от разрядности аналого-цифрового преобразователя (АЦП), типа преобразуемого сигнала и формы шкалы квантователя
- Следует заметить, что при дискретизации сигнала с частотой $f_d > 2f_{гр}$ спектр шума квантования распространяется вплоть до частоты $f_d/2$ и при восстановлении сигнала с помощью ФНЧ с граничной частотой примерно равной $f_{гр}$, уровень шума может быть уменьшен
- При равномерной дискретизации сигнала с числом уровней квантования 2^N отношение сигнал/шум (динамический диапазон АЦП) оценивают в дБ по формуле (C – константа, зависящая от формы преобразуемого сигнала):

$$S = 20N \lg 2 + 10 \lg \frac{f_d}{2f_{гр}} + C \cong 6N + 10 \lg \frac{f_d}{2f_{гр}} + C$$

- При преобразовании синусоидального сигнала $C = 1,7$ дБ, для звуковых сигналов константа C изменяется от -15 дБ до 2 дБ

- Принято, что в звуковой аппаратуре максимально допустимый уровень сигнала соответствует 0 дБ. В связи с этим величина $-S$ характеризует уровень шумов квантования и одновременно минимально возможный уровень полезного сигнала. Из приведенного выше соотношения следует, что при $f_d = 4f_{гр}$ интенсивность шума уменьшится на 3 дБ
- На рисунке приведена диаграмма уровней сигнала при аналого-цифровом преобразовании. Очевидно, что отсчеты сигнала не должны превышать некоторого заданного максимально допустимого уровня. В противном случае квантователь «обрезает» сигнал. Это явление называют клиппингом (англ. clipping). На рисунке иллюстрируется также формирование шума квантования. При цифровом преобразовании аналогового сигнала $u(t)$ различия между его истинными значениями в точках дискретизации и цифровыми отсчетами и определяют шум квантования. Специфическим видом шума квантования является **гранулярный шум** (granular noise), проявляющийся в следствие неустойчивости операции округления, когда уровень аудиосигнала незначительно изменяется и располагается приблизительно посередине между двумя ближайшими уровнями квантования j и $j+1$



- Для оцифровки речевой информации, ограниченной по спектру до 2-5 кГц, обычно используют 7- или 8-разрядные АЦП при частотах дискретизации от 8 до 10 кГц
- В современных цифровых системах обработки и кодирования звуковой информации применяют 16-разрядное квантование и стандартные частоты дискретизации 44,1 или 48 кГц, ограничивая частотный диапазон сигнала примерно равным 20 кГц
- В студийной аппаратуре зачастую используют 18-, 20-, 24- и 32-разрядное квантование при частотах дискретизации 56, 96 и 192 кГц, что обеспечивает сохранность высших гармоник звукового сигнала, непосредственно не воспринимаемых слухом, но влияющих на формирование общей звуковой картины. При ограничении спектра сигнала, учитывая что $f_d > 2f_{гр}$, дополнительно возможно уменьшить уровень шумов квантования
- В таблице приведены характеристики цифрового звукового сигнала для различных видов аналого-цифрового преобразования (при расчете отношения сигнала к шуму квантования не учтено влияние константы C , величина которой зависит от формы преобразуемого сигнала)

Разрядность АЦП, N	Частота дискретизации, кГц	Цифровой поток, кбит/с	Отношение сигнал/шум $6N$, дБ	Отношение сигнал/шум при ограничении спектра сигнала до частоты $f_{гр}$
7 – 8	8 – 12	56 – 96	42 – 48	-
16	44,1	705,6	96	96
16	48	768	96	96,4
18	56	1008	108	109
20	96	1920	120	123,4
24	96	2304	144	147,4
24	192	4608	144	148,4
32	192	6144	192	196,4

- В таблице приведены параметры цифрового потока импульсно-кодовой модуляции (ИКМ) монофонического сигнала. При преобразовании стереофонического или квадрофонического звукового сигнала величина цифрового потока должна быть увеличена в два или четыре раза соответственно
- Очевидно, чтобы влияние шумов квантования было не очень большим, необходимо максимально использовать весь динамический диапазон АЦП, поддерживая уровень преобразуемого аналогового сигнала в определенных границах. Этот процесс реализуется путем использования компараторов, обеспечивающих обработку динамично изменяющегося уровня звукового сигнала. Однако этот метод не всегда приводит к положительным результатам, особенно в тех случаях, когда в течение некоторого времени звуковая информация попеременно изменяется от тихой к громкой. В этих случаях прибегают к иным методам уменьшения влияния шума квантования
- Один из них, называемый **дизерингом** (от англ. «dithering» - «дрожание») используется в тех случаях, когда шум существенно зависит от преобразуемого сигнала и проявляется на слух в виде специфической мешающей помехи (что практически всегда имеет место). Принцип дизеринга заключается в добавлении к преобразованному сигналу псевдослучайной последовательности малого по уровню (не более двух дискретных значений) шума, реализуя тем самым практически полную независимость шума от исходного сигнала, что для слуха является более приемлемым
- Другой принцип снижения влияния связанных с формой сигнала шумов квантования, называемый методом **формовки шума** (noise shaping), заключается в преобразовании спектра шума таким образом, чтобы большая часть его энергии располагалась в менее заметных для слуха областях спектра в соответствии с формой кривых равной громкости

- Применение рассмотренных методов преобразования шумов квантования не всегда рационально, особенно в случаях необходимости дальнейшей обработки аудиосигналов
- Поэтому чаще всего для уменьшения влияния шумов квантования обработку аудиосигналов производят с применением АЦП с большей разрядностью при более высокой частоте дискретизации
- При этом чем выше частота дискретизации, тем большие требования предъявляются к ее стабильности, поскольку несовершенства преобразующей аппаратуры, приводящие к случайным временным отклонениям дискретизирующих импульсов вызывают эффект **джиттера** (от англ. «jitter» - «дрожание»)
- Для борьбы с джиттером применяются высокостабильные кварцевые генераторы

Аналого-цифровое и цифро-аналоговое преобразование звуковых сигналов

- Кодирующее и декодирующее устройства на входе и выходе оперируют с аналоговыми сигналами, преобразование которых в цифровой код и их восстановление являются важнейшими этапами обработки информации. Сигналы на выходе декодера даже при отсутствии линейной и нелинейной обработки информации практически никогда не совпадают по форме с сигналами, поступающими на вход кодера. Ниже приводятся наиболее широко используемые варианты преобразования звуковых сигналов:
- Импульсно-кодовая модуляция (ИКМ)
- Дифференциальная импульсно-кодовая модуляция (ДИКМ)
 - дельта-модуляция
 - адаптивная дельта-модуляция
 - адаптивная относительная импульсно-кодовая модуляция
- Сигма-дельта модуляция (СДМ)

Импульсно-кодовая модуляция (ИКМ)

- ИКМ – наиболее простой способ преобразования сигналов, обычно содержащий в кодере многоразрядный АЦП (чаще с линейной шкалой квантования) и в декодере ЦАП, имеющий такую же разрядность
- Если число N – разрядность АЦП, а частота дискретизации сигнала равна f_δ , то формируемый цифровой поток определяется их произведением $N \cdot f_\delta$ кбит/с
- Идеализированное восстановление аналогового сигнала представляет собой последовательность импульсов, но реальные ЦАП создают «ступенчатую» форму сигнала, применяя к выходным отсчетам удержание нулевого порядка. В частотной области удержание нулевого порядка приводит к выходному спектру, который соответствует идеализированному спектру, умноженному на функцию sinc.
- На выходе ЦАП при этом последовательно с частотой дискретизации f_δ выделяются импульсные дискретные значения звукового сигнала. Каждый из этих почти прямоугольных импульсов длительностью $\tau = 1/f_\delta$ обладает sinc-спектром, плавно изменяющимся от 1 на частоте $f = 0$ до $2/\pi$ на частоте $f_\delta/2$

$$F(f) \equiv \sin c \pi \cdot f / f_\delta = \frac{\sin \pi \cdot f / f_\delta}{\pi \cdot f / f_\delta}$$

- Если частота дискретизации мало отличается от удвоенной величины граничной частоты преобразуемого сигнала $f_\delta \approx 2 f_{гр}$, то спектральные составляющие сигнала в области граничной частоты оказываются значительно ослабленными. Устранение этого эффекта возможно либо путем использования ФНЧ с соответствующим подъемом частотной характеристики в области граничной частоты сигнала, либо с применением sinc-предкоррекции дискретных отсчетов АЦП

Sinc-предкоррекция

Форма частотной характеристики предкорректирующей цепи определяется формулой:

$$F_{np}(f) = \frac{\pi \cdot f / f_{\partial}}{\sin \pi \cdot f / f_{\partial}} = a_0 + 2 \sum_{i=1}^{\infty} a_i \cos(2i\pi \cdot f / f_{\partial})$$

где
$$a_0 = \frac{4}{\pi} G, \quad a_i = \frac{4}{\pi} \left[G - \sum_{k=0}^{i-1} \frac{(-1)^k}{(2k+1)^2} \right],$$

$$G = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} = 0,915965594... \quad - \text{постоянная Каталана}$$

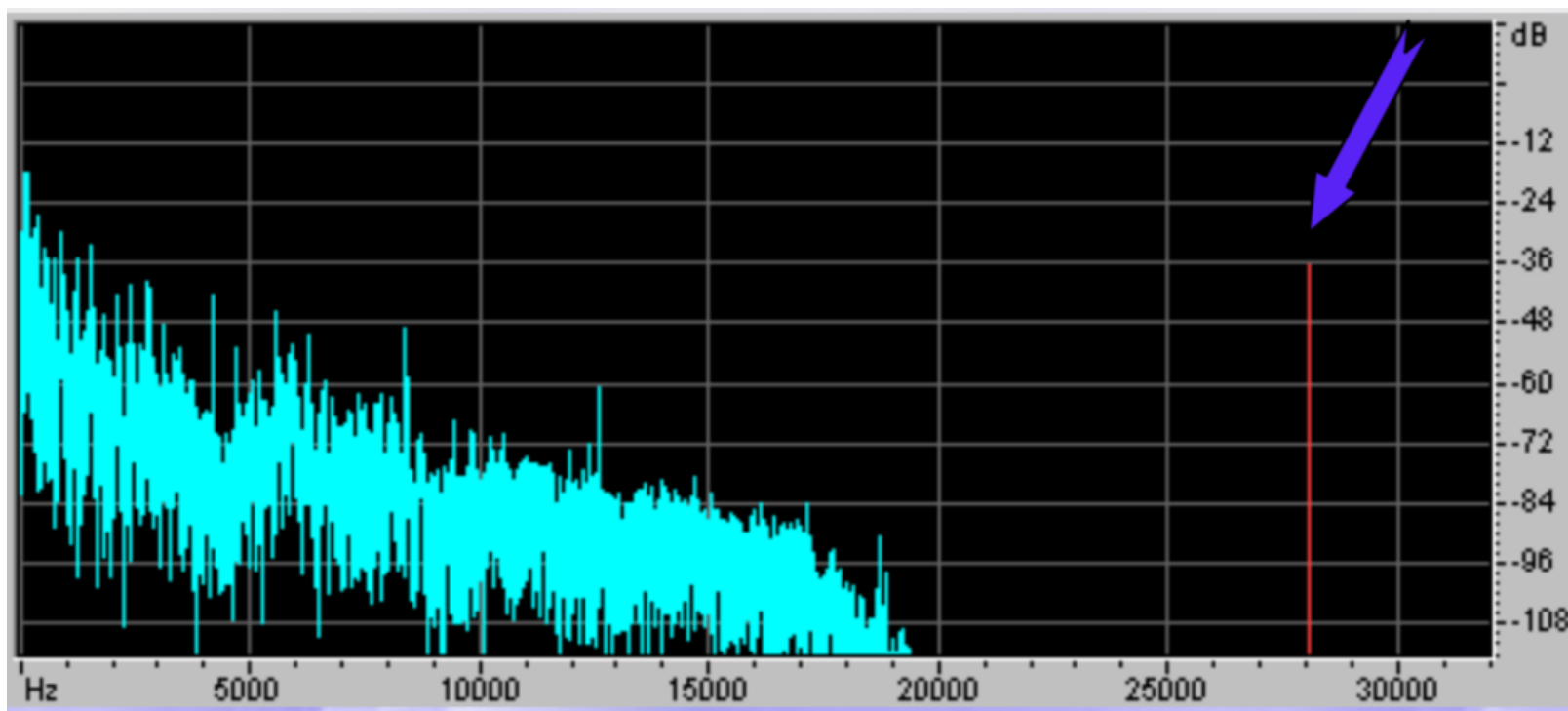
Ряд достаточно быстро сходится, по этой причине можно ограничиться конечным числом его членов. С точностью выше 0,1 % процесс предварительной коррекции реализуется в соответствии с соотношением:

$$\bar{u}(n) = a_0 \cdot u(n) + \sum_{k=1}^4 a_k \cdot [u(n-k) + u(n+k)]$$

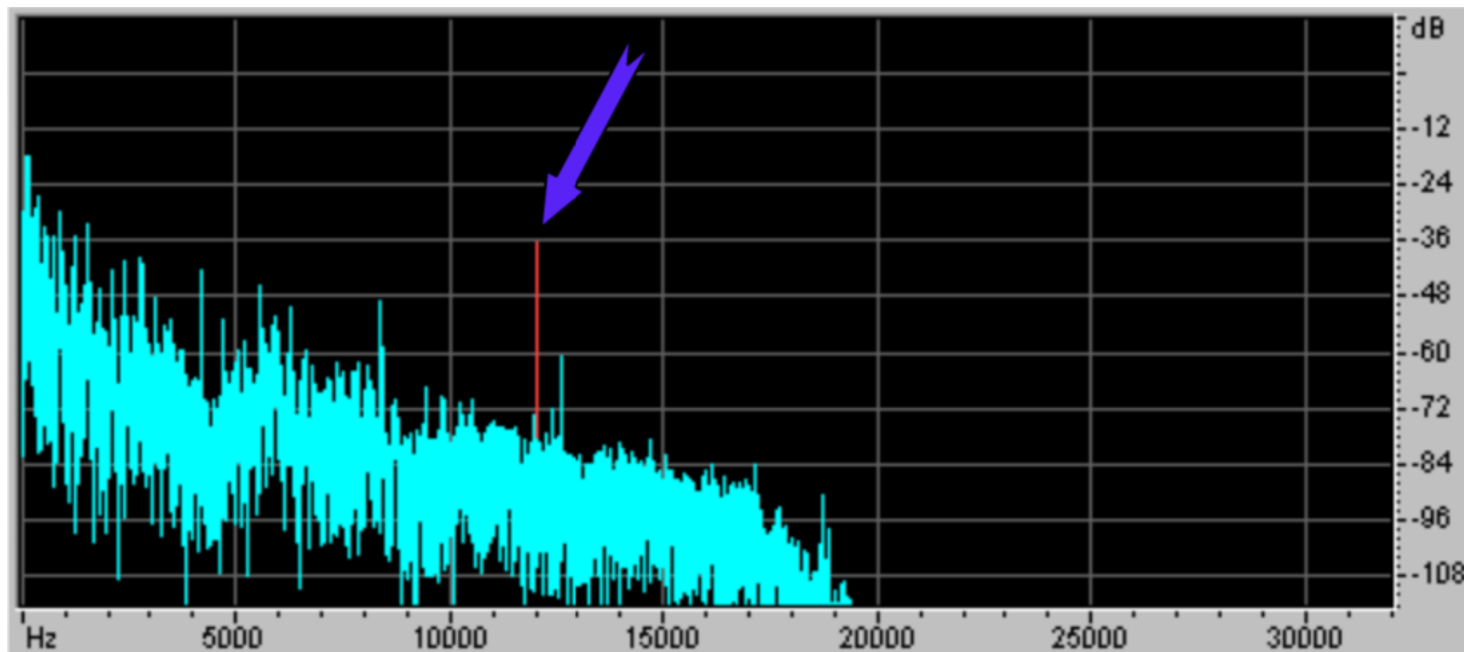
где $a_0 = 1,1662$, $a_1 = -0,1070$, $a_2 = 0,0345$, $a_3 = -0,0165$, $a_4 = 0,0095$.

Алиасинг

- Пусть звук не содержит частот выше 20 кГц. Тогда, по теореме Котельникова, можно выбрать частоту дискретизации 40 кГц.
- Пусть в звуке появилась помеха с частотой 28 кГц. Условия теоремы Котельникова перестали выполняться.



- Проведем дискретизацию с частотой 40 кГц, а затем – восстановим аналоговый сигнал sinc-интерполяцией



- Помеха отразилась от половины частоты дискретизации в нижнюю часть спектра и наложилась на звук. Помеха переместилась в слышимый диапазон – это алиасинг.

- Как избежать алиасинга? Применить перед оцифровкой анти-алиасинговый фильтр. Он подавит все помехи выше половины частоты дискретизации (выше 20 кГц) и пропустит весь сигнал ниже 20 кГц.
- После этого условия теоремы Котельникова будут выполняться и алиасинга не возникнет.
Следовательно, по цифровому сигналу можно будет восстановить исходный аналоговый сигнал.

Дифференциальная импульсно-кодовая модуляция (ДИКМ)

- На вход квантователя последовательно подаются отсчеты не исходного сигнала, а разность между его текущим значением и предсказанной величиной $e(n) = u(n) - \tilde{u}(n)$
- При использовании метода линейного предсказания величина $\tilde{u}(n)$ определяется соотношением

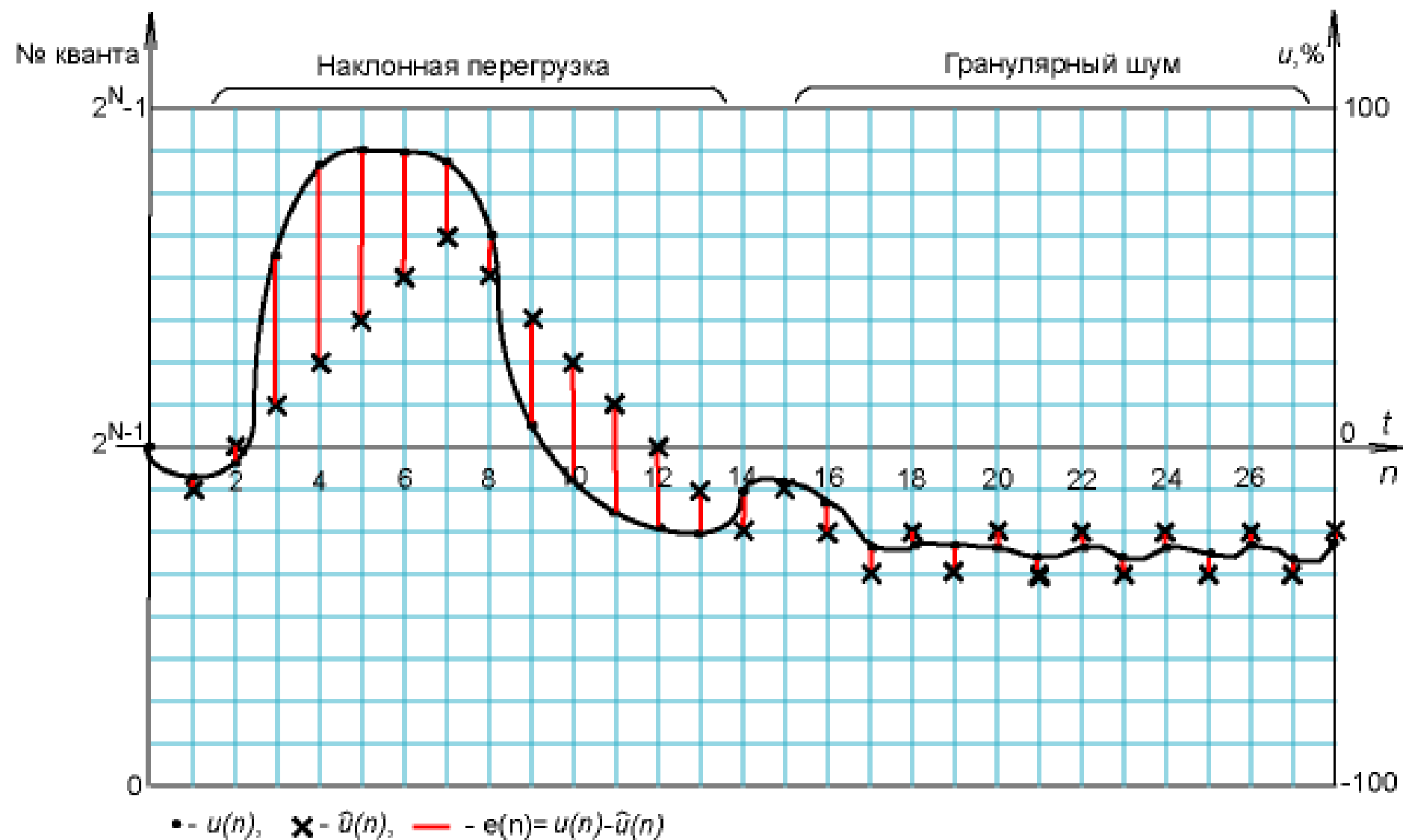
$$\tilde{u}(n) = \sum_{k=1}^K a_k \hat{u}(n-k)$$
 где K и a_k – порядок и коэффициенты предсказания, $\hat{u}(n-k)$ – отсчеты сигнала, полученные на предыдущих шагах преобразования исходных величин $u(n-k)$
- Формирование величин $\hat{u}(n)$ определяется принципом кодирования информации о разностной величине $e(n)$ и с учетом поправки $Q[e(n)]$

$$\hat{u}(n) = \tilde{u}(n) + Q[e(n)]$$
- Величины коэффициентов предсказания зависят от корреляционных свойств преобразуемой аудиоинформации и подбираются опытным путем. Например, при оцифровке речевой информации может быть использован предсказатель пятого порядка с коэффициентами $a_1=0,86$; $a_2=0,64$; $a_3=0,40$; $a_4=0,26$; $a_5=0,20$

- Наиболее простым способом кодирования с предсказанием является **дельта-модуляция (ДМ)**, реализуемая с помощью однобитного квантователя. Ошибка предсказания $e(n)$ и поправка $Q[e(n)]$ определяются соотношениями:

$$e(n) = u(n) - \hat{u}(n-1); \quad Q[e(n)] = \begin{cases} 1, & \text{при } e(n) \geq 0; \\ 0, & \text{при } e(n) < 0. \end{cases}$$

- Принцип формирования сигнала на выходе декодирующего устройства заключается в следующем:
 - если величина $Q[e(n)]=1$, то значение отчета выходного сигнала вычисляется по формуле $\hat{u}(n) = \hat{u}(n-1) + \Delta$, где Δ – некоторое дискретное приращение сигнала;
 - при $Q[e(n)]=0$ результат преобразования отсчета имеет вид $\hat{u}(n) = \hat{u}(n-1) - \Delta$
- Применение такого вида преобразования приводит к двум видам искажений – **перегрузке крутизны** (или наклонной перегрузке), связанной с неспособностью кодирующего устройства реагировать на быстрые изменения величин отсчетов аудиосигнала, и появлению **гранулярного шума**. В связи с этим использование дельта-модуляции эффективно при высокой корреляции соседних отсчетов сигнала и применяется при повышенных частотах дискретизации звуковых сигналов



- Более эффективно использование **адаптивной дельта-модуляции (АДМ)**, при которой в зависимости от характера поступающих на вход кодирующего устройства отсчетов сигнала изменяется шаг квантования
- Наиболее простой способ изменения шага квантования поясняется следующим примером

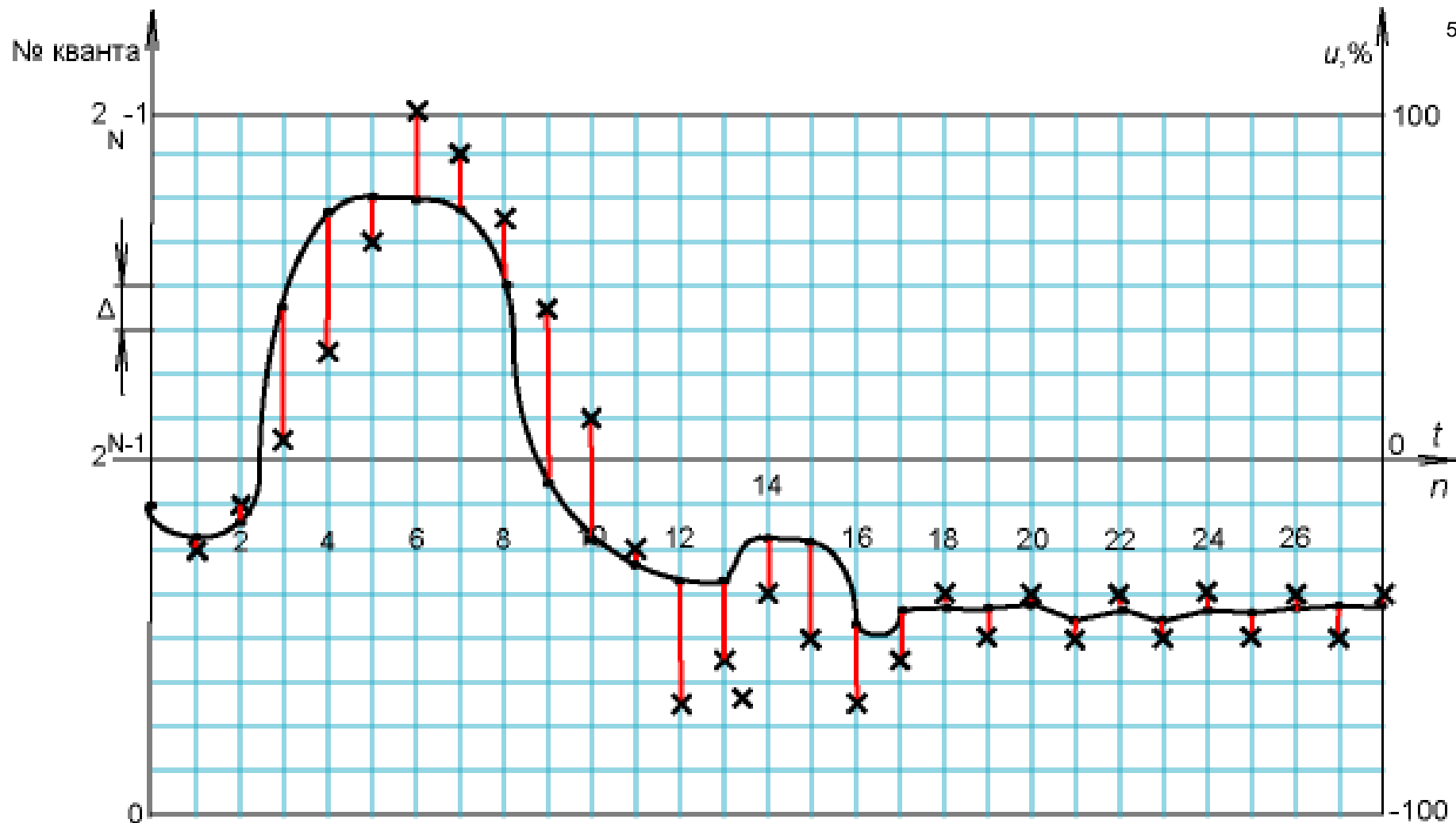
На начальном этапе преобразования в качестве исходного шага квантования принимаются значение поправки $Q[e(n)]=1$ и величина шага квантования $\Delta(0)=\Delta$

Последующее изменение шага квантования осуществляется, например, в соответствии с формулой:

$$\Delta(n) = \begin{cases} \Delta(n-1) + \Delta/2, & \text{если } Q[e(n)] = Q[e(n-1)]; \\ \Delta(n-1) - \Delta/2, & \text{если } Q[e(n)] \neq Q[e(n-1)]. \end{cases}$$

Кроме того, $\Delta(n) \geq \Delta/2$

- Рисунок иллюстрирует более эффективное кодирование быстро изменяющегося сигнала и частичную компенсацию нежелательного эффекта перегрузки крутизны. Гранулярный шум также компенсируется за счет автоматического уменьшения шага квантования на каждом новом этапе преобразования



Значение $Q[e(n)]$ на выходе квантователя:

1 0 1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 1 1 0 1 0 1 0 1 0 1 0 1

Величина шага квантования $\Delta(n) = k\Delta$, где k :

1 $\frac{1}{2}$ $\frac{1}{2}$ 1 $\frac{3}{2}$ 2 $\frac{5}{2}$ $\frac{1}{2}$ 1 $\frac{3}{2}$ 2 $\frac{5}{2}$ 3 $\frac{1}{2}$ 1 $\frac{1}{2}$ 1 $\frac{1}{2}$ 1 $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$

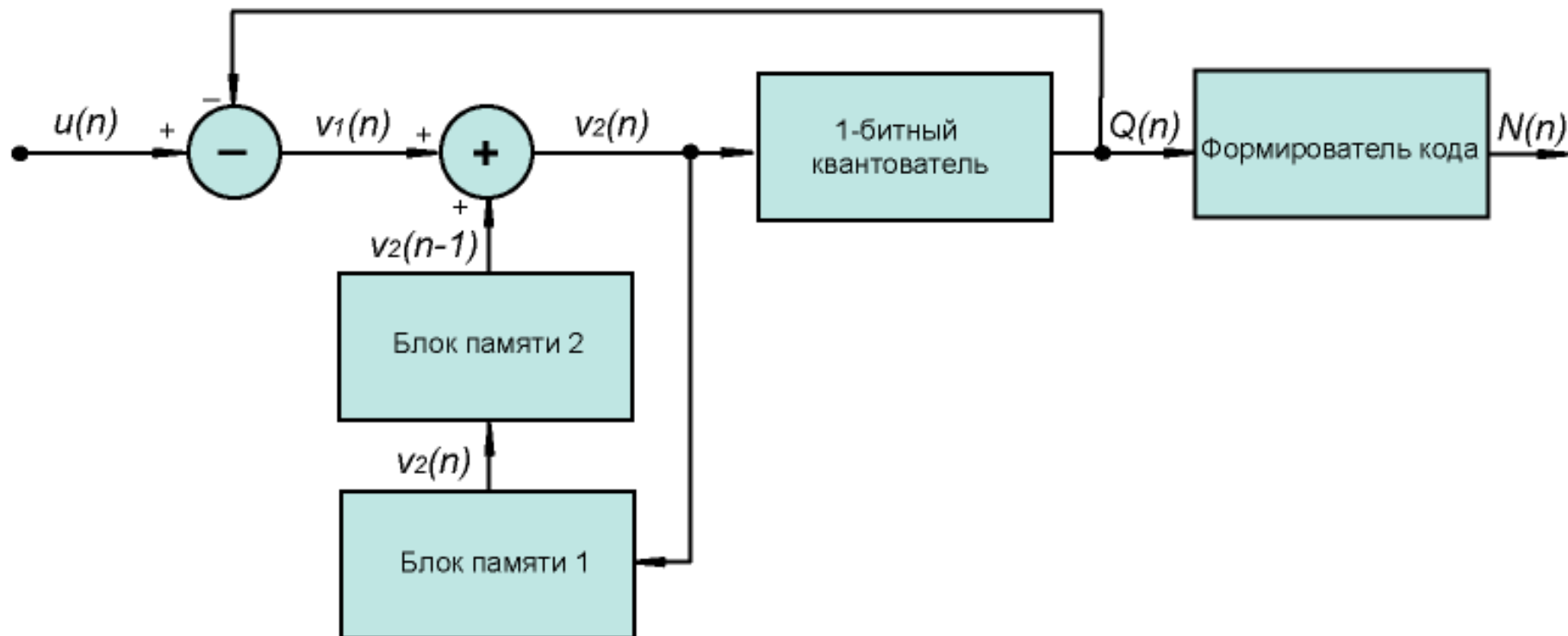
• - $u(n)$, **x** - $\hat{u}(n)$, — - $e(n) = u(n) - \hat{u}(n)$

- Еще большая эффективность кодирования аудиоинформации может быть получена при квантователе, использующем предсказание более высокого порядка и реализуемом **адаптивной относительной импульсно-кодовой модуляции АОИКМ**
- **Пример одно из вариантов АОИКМ:**
 Шаг квантования $\Delta(n)$ зависит от результата кодирования на предыдущем шаге $\Delta(n) = M \cdot \Delta(n-1)$, а функция квантования $Q[e(n)]$ зависит от величины $\Delta(n)$ и передается на декодер тремя битами
 Значение коэффициента M , увеличивающего или уменьшающего шаг квантования определяется следующим соотношением:

$$M = \begin{cases} 0,5 & \text{при} & 0 \leq |e(n-1)| \leq \Delta(n-1); \\ 1,5 & \text{при} & \Delta(n-1) < |e(n-1)| \leq 2\Delta(n-1); \\ 2,5 & \text{при} & 2\Delta(n-1) < |e(n-1)| \leq 3\Delta(n-1); \\ 3,5 & \text{при} & 3\Delta(n-1) < |e(n-1)|. \end{cases}$$
- Данная схема эффективна, весьма проста и легко реализуется в цифровом виде. Существуют более сложные и более эффективные схемы реализации АОИКМ с многобитной функцией квантования $Q[e(n)]$

Сигма-дельта модуляция (СДМ)

- Основой сигма-дельта модуляции является не анализ приращений сигнала, а кодирование уровней самого преобразуемого сигнала, как при ИКМ
- Структурная схема сигма-дельта модулятора приведена на рисунке
- На вход модулятора подаются дискретные отсчеты сигнала $u(n)$, относительные уровни которых могут изменяться в пределах от -1 до $+1$



- Кодер содержит однобитовый квантователь, на выходе которого формируется сигнал:

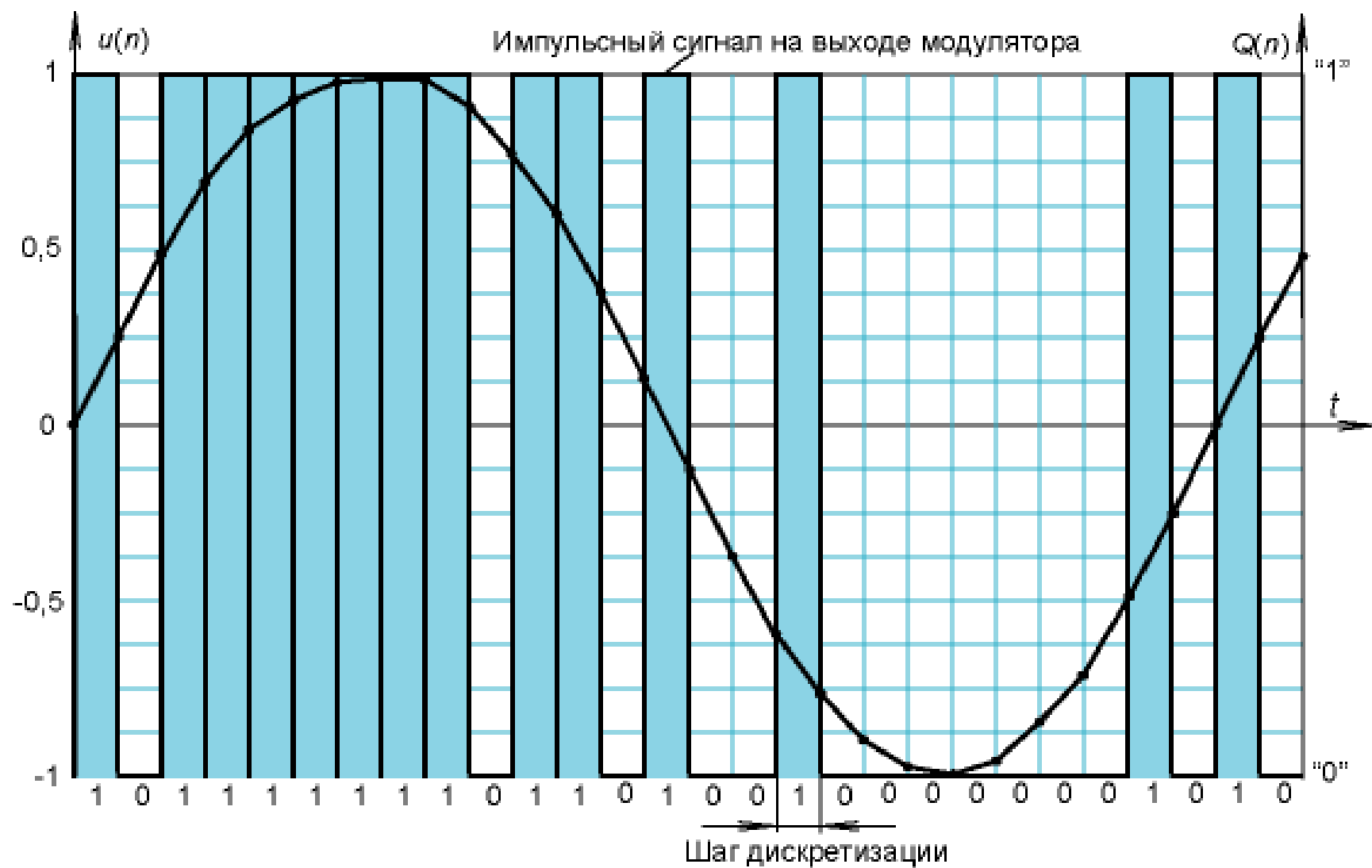
$$Q(n) = \begin{cases} +1, & \text{если } v_2(n) \geq 0; \\ -1, & \text{если } v_2(n) < 0, \end{cases}$$

где $v_2(n)$ – дискретные отсчеты сигнала на входе квантователя, формируемые следующим образом:

- вычитатель, включенный на входе устройства обеспечивает формирование разности отсчетов входного сигнала $u(n)$ и выходного сигнала квантователя $Q(n)$: $v_1(n) = u(n) - Q(n)$;
- отсчеты разностного сигнала $v_1(n)$ подаются на сумматор с выхода которого отсчеты сигнала $v_2(n)$ поступают параллельно на входы квантователя и блока памяти БП1;
- с выхода блока БП1 с задержкой на такт отсчеты сигнала подаются на вход блока памяти БП2, на выходе которого выделяются отсчеты сигнала предыдущего такта $v_2(n - 1)$, которые поступают на второй вход сумматора;
- таким образом, на вход квантователя подается сигнал $v_2(n) = v_1(n) + v_2(n - 1)$.
- Формирователь кода, включенный на выходе квантователя, создает цифровую последовательность в соответствии с соотношением:

$$N(n) = \begin{cases} 1, & \text{если } Q(n) = 1; \\ 0, & \text{если } Q(n) = -1. \end{cases}$$

№ такта	Входной сигнал $u(n)$	На выходе вычитателя $v_1(n)=u(n)-Q(n)$	На входе БП2 $v_2(n-1)$	На выходах слагателя и БП1 $v_2(n)=v_1(n)+v_2(n-1)$	На выходе квантователя $Q(n)$
0	0,00	0,00	0,00	0,00	1
1	0,25	-0,75	0,00	-0,75	-1
2	0,48	1,48	-0,75	0,73	1
3	0,68	-0,32	0,73	0,41	1
4	0,84	-0,16	0,41	0,25	1
5	0,95	-0,05	0,25	0,20	1
6	1,00	0,00	0,20	0,20	1
7	0,98	-0,02	0,20	0,18	1
8	0,90	-0,10	0,18	0,08	1
9	0,77	-0,23	0,08	-0,15	-1
10	0,59	1,59	-0,15	1,44	1
11	0,37	-0,63	1,44	0,81	1
12	0,13	-0,87	0,81	-0,06	-1
13	-0,13	0,87	-0,06	0,81	1
14	-0,37	-1,37	0,81	-0,56	-1
15	-0,59	0,41	-0,56	-0,15	-1
16	-0,77	0,23	-0,15	0,08	1
17	-0,90	-1,90	0,08	-1,82	-1
18	-0,98	0,02	-1,82	-1,80	-1
19	-1,00	0	-1,80	-1,80	-1
20	-0,95	0,05	-1,80	-1,75	-1
21	-0,84	0,16	-1,75	-1,59	-1
22	-0,68	0,32	-1,59	-1,27	-1
23	-0,48	0,52	-1,27	-0,75	-1
24	-0,25	0,75	-0,75	0,00	1
25	0,00	-1,00	0,00	-1,00	-1
26	0,25	1,25	-1,00	0,25	1
27	0,48	-0,52	0,25	-0,27	-1



- Преимущества СДМ:
 - Простая техническая реализация (по сравнению с ИКМ с линейным многобитным квантователем)
 - Вследствие применения более высоких частот дискретизации, по сравнению с ИКМ, шум квантования простирается на более широкую полосу
44,1 кГц / 16 бит – шум в полосе до 22,05 кГц
СДМ 705,6 кГц – шум в полосе до 352,8 кГц ($44,1 \times 16 = 705,6$)
- Недостатки СДМ
 - Высокая частота дискретизации для достижения высокого качества кодирования
- СДМ также называют модуляцией плотностью импульсов (PDM – pulse density modulation)
- Формат СДМ легко может быть преобразован в формат ИКМ
- Стандарт SADC (Super Audio CD) (Sony, Phillips, 1997): частота дискретизации – 2,8224 МГц на канал

Лабораторная работа по теме «Основы обработки звука в MATLAB»

Задача 1: анализ зашумленного звукового файла.

Дано: аудиофайл длиной 1 с с тремя поврежденными гармониками неизвестной частоты.

- 1) Найти частоту гармоник (0,5 балла).
- 2) Отфильтровать аудиофайл так, чтобы снизить помехи (1 балл).

Задача 2: анализ клиппинг-эффекта.

- 1) Создайте звуковой файл из синусоиды с частотой 1 КГц, амплитудой 3 и длиной 3с с частотой дискретизации 2500 Гц. (0,5 балла).
- 2) Сымитируйте клиппинг-эффект (пусть ваш детектор имеет максимум в 2 вольты). Проанализируйте, как это отразится на характеристиках сигнала (1 балл).

Задача 3: анализ влияния частоты дискретизации.

Дано: аудиофайл.

- 1) Найдите частоту дискретизации. Уменьшите её вдвое. Проанализируйте, как это отразится на субъективном качестве и характеристиках сигнала (0,5 балла).
- 2) Удвойте частоту дискретизации. Недостающие отсчеты заполните средним соседних отсчетов (1 балл).

Дополнительная задача: 1,5 балла.

Дан аудиофайл с записью звуков волнистого попугайчика. В состоянии раздражения они издают характерный звук «ч». Маскируйте этот звук (он должен остаться в итоговом аудиофайле, но не должен быть слышен).