

2.2. Особенности представления и восприятия звука

Звук

2

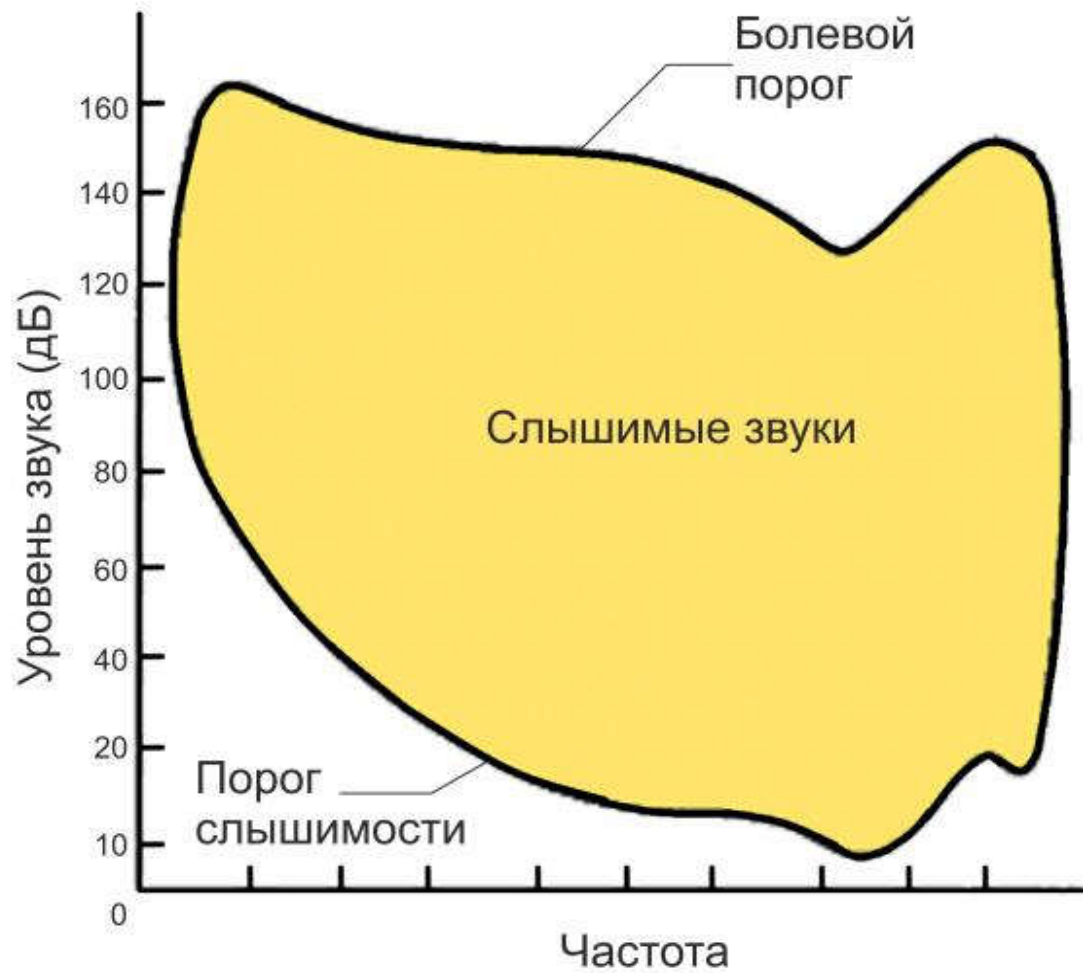
- Звук — это упругая волна, продольно распространяющаяся в среде и создающая в ней механические колебания.
- Меняющейся (колеблющейся) характеристикой является давление в точке среды
- Зависимость длины волны λ от частоты имеет вид $\lambda = \frac{u}{\nu}$
- Слуховая система человека различает звуки
 - частоты 20 ... 20000 Гц,
 - давлением 10 мкПа ... 100 Па
- Для удобства сравнения волн в столь широком диапазоне амплитуд применяется логарифмическая мера:

$$l_{1,2} = 20 \lg \frac{p_1}{p_2}$$

- $l = 20 \lg \frac{p}{p_0}$ — уровень звука, где $p_0 = 20,4$ мкПа

Область слышимых звуков

3



Цифровой звук

4

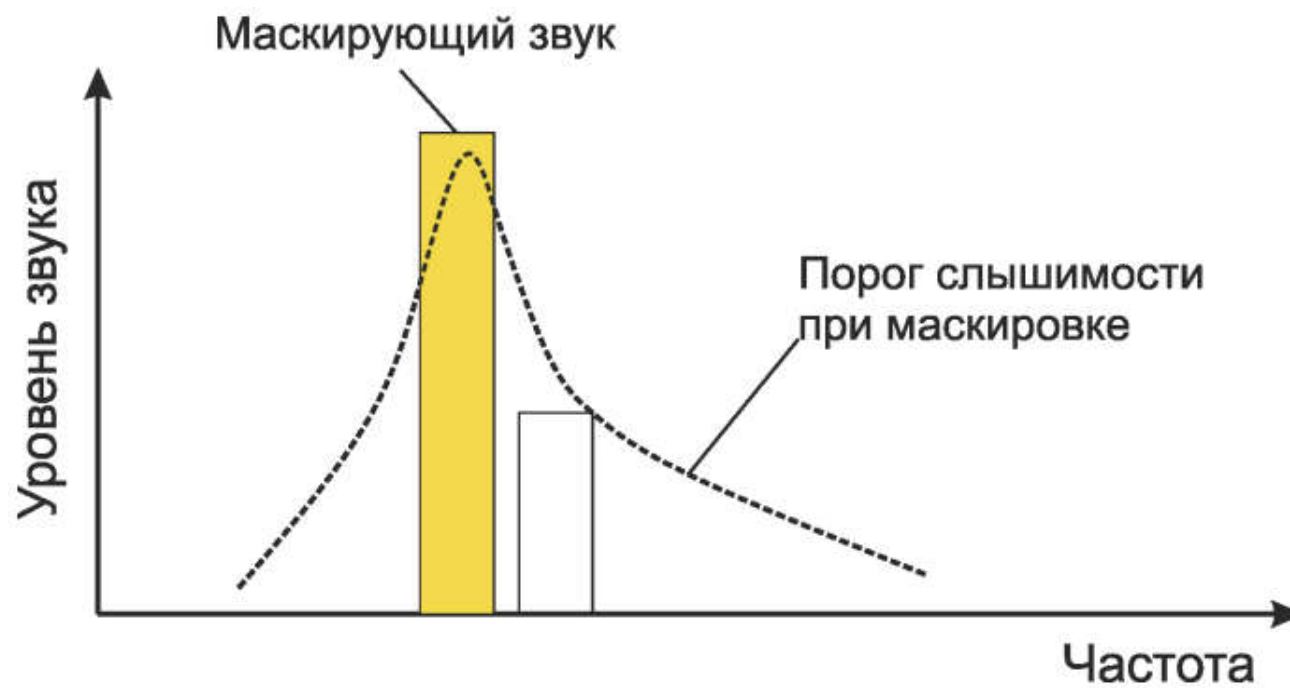
- В целом слуховая система человека настроена на восприятие человеческой речи в диапазоне 50–5000 Гц.
- Аудиосигналы можно разделить на три класса:
 - ▣ разговор телефонного качества, диапазон 300–3400 Гц;
 - ▣ широкополосная речь 50–7000 Гц;
 - ▣ широкополосные аудиосигналы 20–20000 Гц.
- При квантовании используется 2^{16} уровней для одного канала
- Стандартная частота дискретизации звукового сигнала составляет 44100 Гц
- *Любая функция, допускающая преобразование Фурье и имеющая непрерывный спектр, ограниченный полосой частот от 0 до ν , полностью определяется дискретным рядом своих значений, отсчитанных через интервалы времени $\Delta t = \frac{1}{2\nu}$*

Продолжительность стереозвука (двухканального), который может быть записан без сжатия на стандартном CD типа «700 МБ»
(позволяющем записать 846720000 байт в формате аудио)

$$\frac{846\,720\,000 \text{ байт на } CD \cdot 8 \frac{\text{бит}}{\text{байт}}}{44100 \text{ Гц} \cdot 16 \frac{\text{бит}}{\text{отсчёт}} \cdot 2 \text{ канала}} = 4800 \text{ с} = 80 \text{ мин.}$$

Частотное маскирование

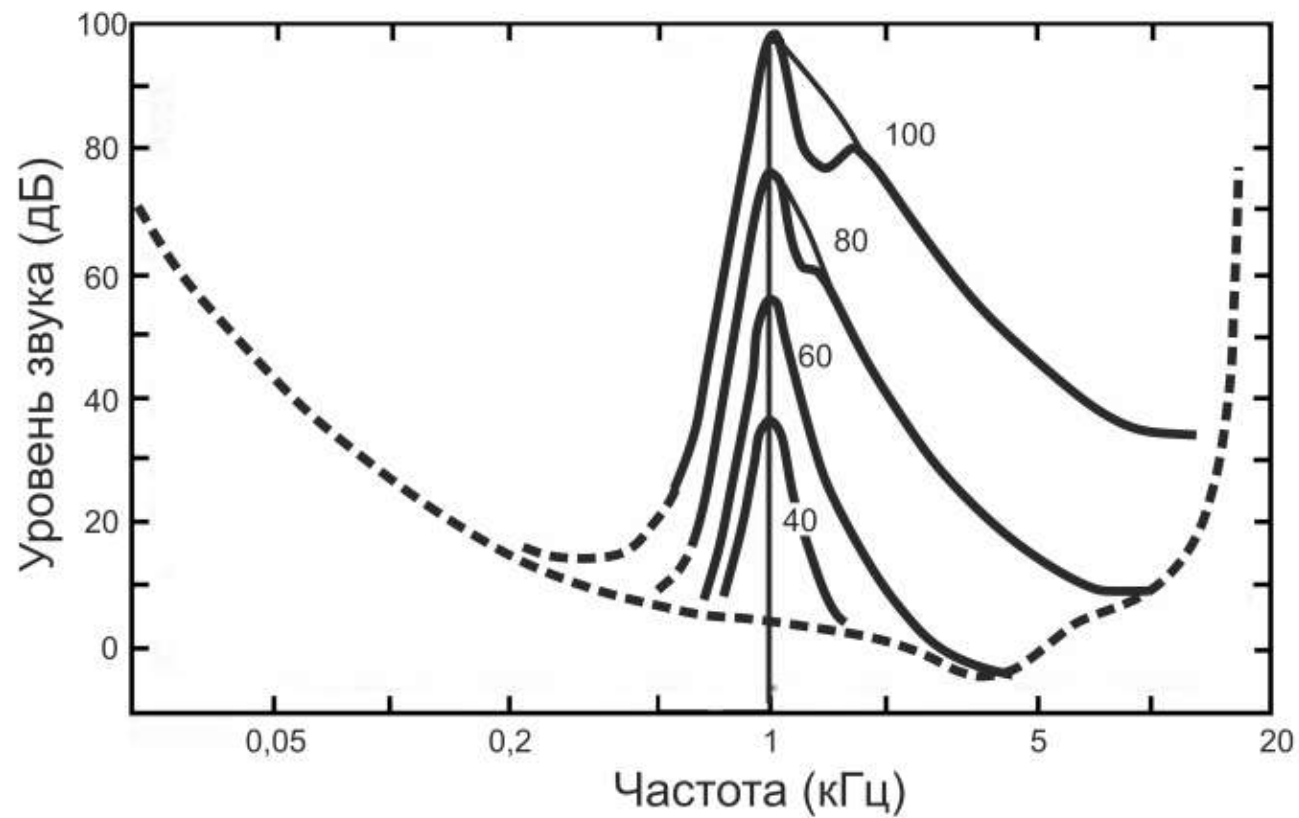
6



Частотное маскирование

7

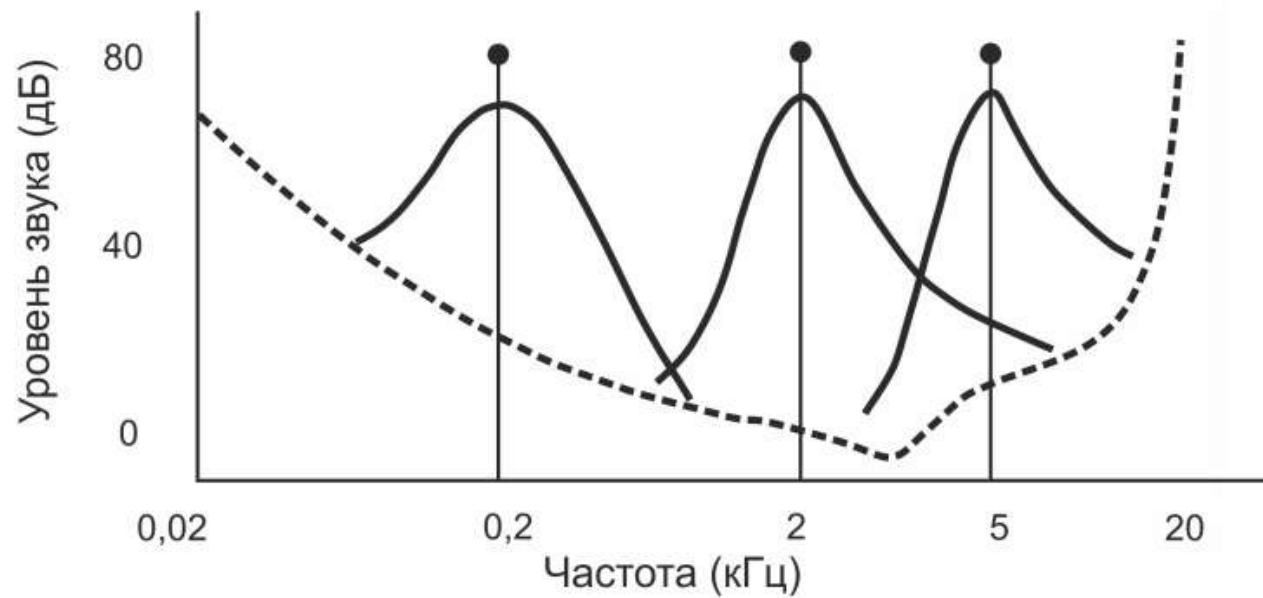
Влияние уровня маскирующего шума на
диапазон маскировки



Частотное маскирование

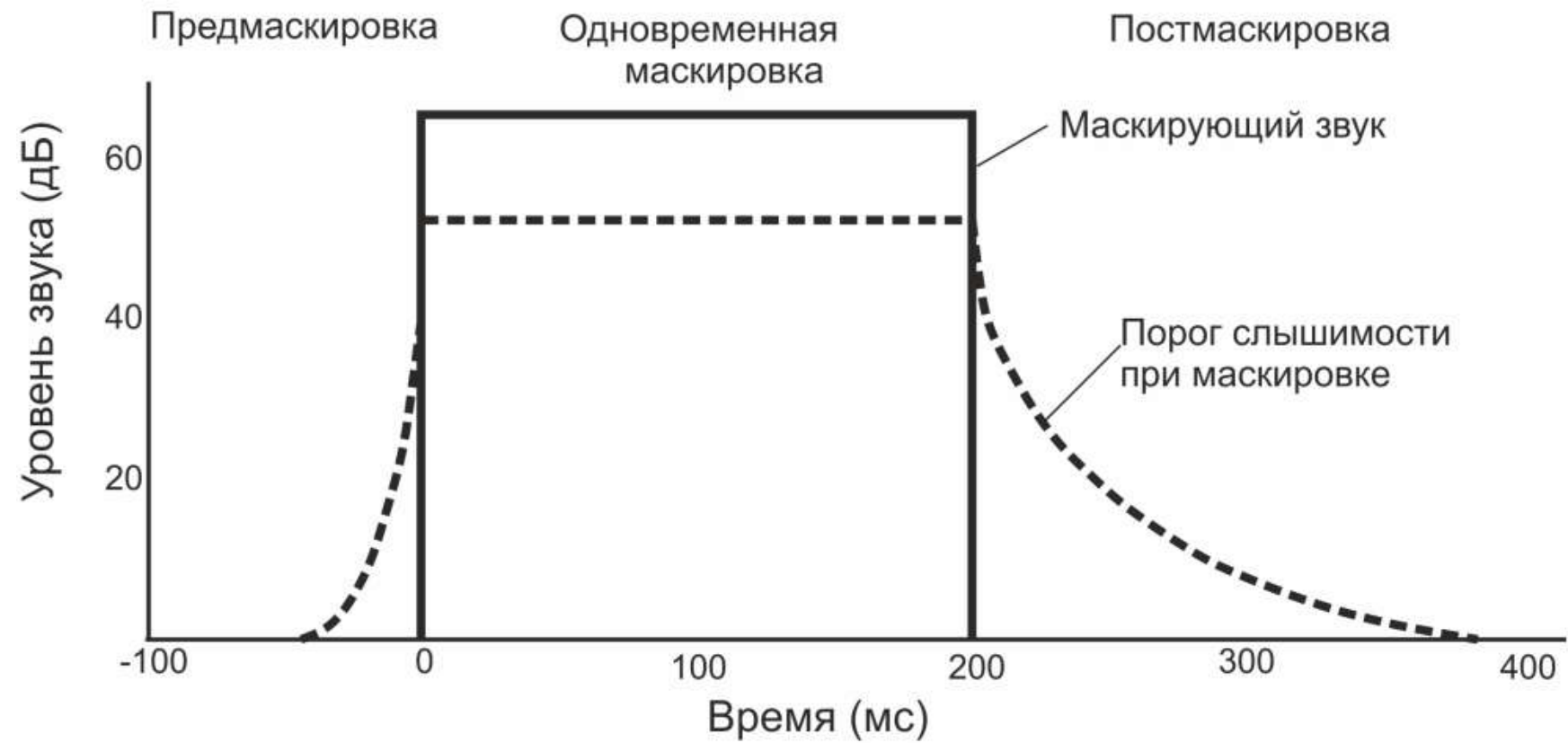
8

Влияние частоты маскирующего шума на
диапазон маскировки



Временная маскировка

9



Показатели качества звуковых сигналов

10

- Искажённый сигнал $v(n)$ длиной N
- Эталонный сигнал $u(n)$ длиной N
- Разностный сигнал $\varepsilon(n) = u(n) - v(n)$

Рассматриваем показатели вида $Q = Q(\varepsilon)$

0. Субъективное аудиальное восприятие

1. Максимальная ошибка

$$\varepsilon_{max} = \max_n |\varepsilon(n)| = \max_n |u(n) - v(n)|$$

2. Среднеквадратичная ошибка

$$\varepsilon_{KB}^2 = \frac{1}{N} \sum_{n=0}^{N-1} \varepsilon^2(n).$$

$$PSNR(u, v) = 10 \cdot \lg \frac{\sup_n^2 u(n)}{\varepsilon_{KB}^2(u, v)} = 20 \cdot \lg \frac{65535}{\varepsilon_{KB}(u, v)}.$$

Частотно-взвешенный среднеквадратичный показатель – трудности

11

Идея: по аналогии с изображениями

$$\varepsilon_{\text{ЧВ КВ}}^2 = \frac{1}{N} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) \cdot |E(e^{i\omega})|^2 d\omega.$$

Не подходит:

1. Звуковой сигнал обладает куда большей длиной N , чем размеры изображения N_1, N_2 .
2. Изображение человек может воспринять целиком за мгновения, в то время как для полноценного восприятия звукового сигнала он должен потратить большое время.

Вывод: делить сигнал на кратковременные фрагменты

Кратковременное преобразование Фурье

12

Кратковременное преобразование Фурье

$$E(m, e^{i\omega}) = \sum_{n=-\infty}^{\infty} \varepsilon(n) \cdot \vartheta(n - m) \cdot e^{-i\omega n},$$

где $\vartheta(n)$ – оконная функция, $\vartheta(n) \neq 0$ для $n \in [0, N_\vartheta - 1]$

Окно Хемминга

$$\vartheta(n) = \alpha + (1 - \alpha) \cos \frac{2\pi n}{N_\vartheta - 1},$$

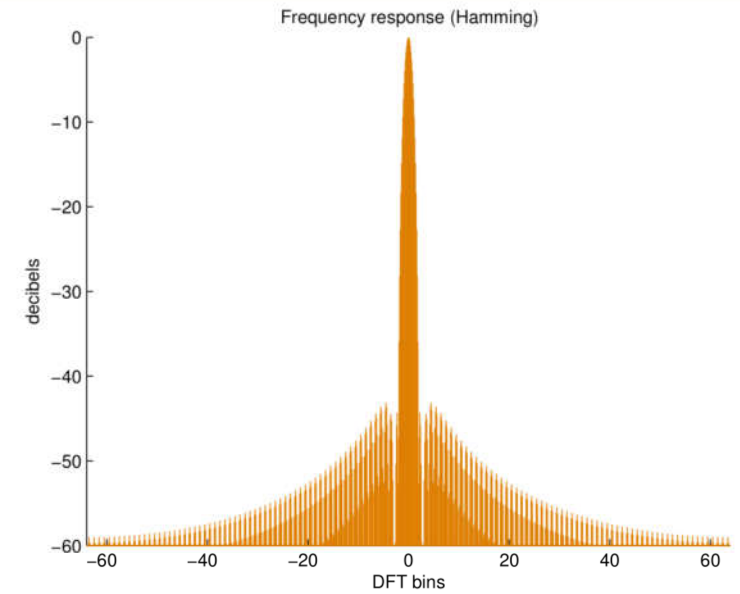
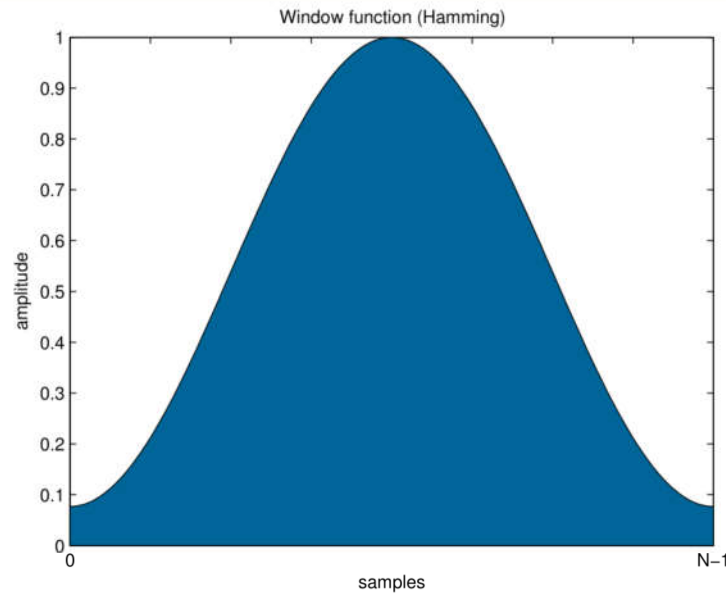
$\alpha = 0,538$ - типичное значение

Случай окна Хемминга при $\alpha = 0,5$ называется окном Ханна.

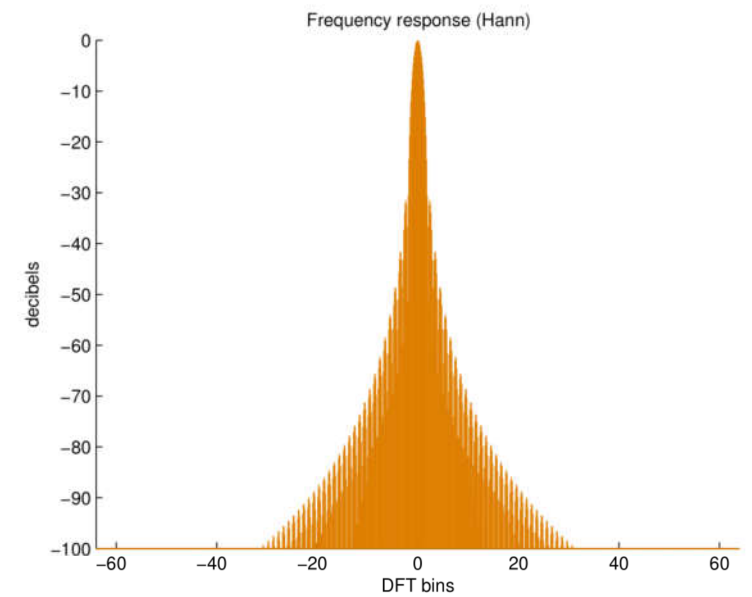
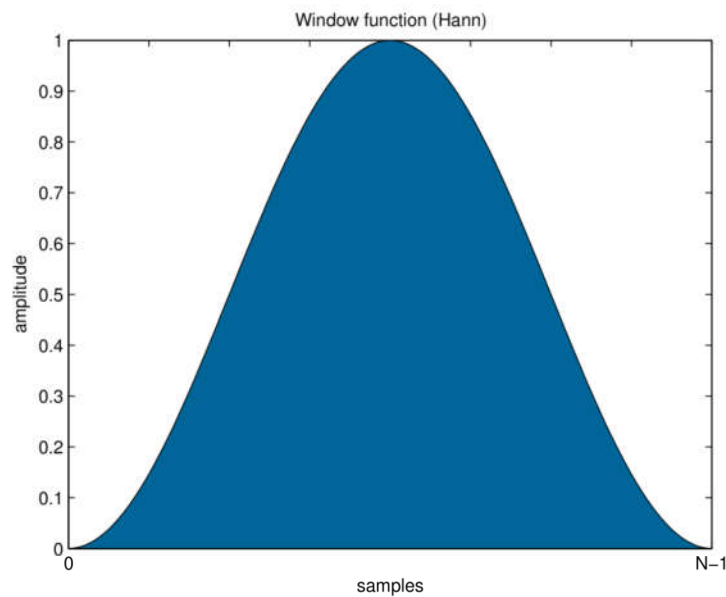
Виды оконных функций

13

□ Окно Хэмминга



□ Окно Ханна



Частотно-взвешенный среднеквадратичный показатель – итог

14

$$\varepsilon_{\text{ЧВ КВ}}^2(u, v) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega) d\omega \cdot \frac{1}{N} \sum_{m=0}^{N-1} |E(m, e^{i\omega})|^2.$$

Свойства $W(\omega)$:

1. $W(\omega) \geq 0$
2. Характеризует чувствительность человеческого слуха в различных частотах.

Решение:

$$W(\omega) = \frac{C}{T(\omega)},$$

где $C > 0$ – коэффициент, $T(\omega)$ – аппроксимация функции пороговой слышимости:

$$T(\omega) = 3,64 \left(\frac{\omega}{1000} \right)^{-0.8} - 6,5 \exp \left\{ -0,6 \left(\frac{\omega}{1000} - 33 \right)^2 \right\} + 10^{-3} \left(\frac{\omega}{1000} \right)^4$$