Лекция 14.
Методы обработки речевых сигналов в задаче распознавания

Докладчик Симончик К.

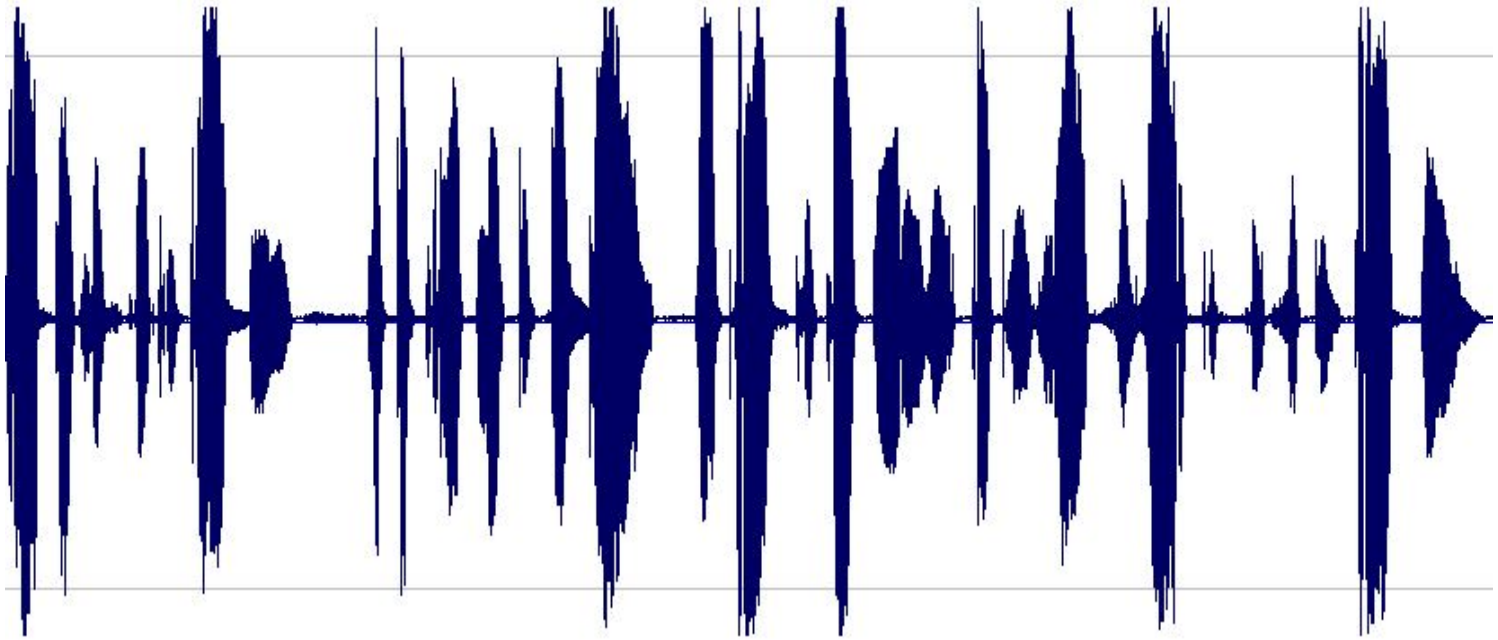
Группа 0382

Содержание

- **Теория сэмплирования**
- **Линейные фильтры**
- **Анализ кратковременного преобразования Фурье**
- **Применение окон**
- **Кодирование речи**

Речевой сигнал

Представление речевого сигнала во временной области



wav-файл, 8000Hz, 16 bit (106 kbyte)

Теория сэмплирования

Частота дискретизации

Перед дискретизацией сигнал необходимо отфильтровать. Теоретически, максимальная воспроизводимая частота является половиной частоты дискретизации

В телефонии использована частота дискретизации 8 кГц. 16 кГц обычно считается достаточным для распознавания и синтеза речи.

Разрешение дискретизации

Уровень звука/ dB SPL	Шумовое отношение	Амплитудное отношение	Типичный пример
120	10^{12}	10^6	Громкая рок-группа
100	10^{10}	10^5	Выстрел в закрытом пространстве
80	10^8	10^4	Оживленная улица
70	10^7	3160	Нормальный разговор
50	10^5	316	Тихий разговор
30	10^3	31.6	Шепот
20	10^2	10	Загородная ночь

Нормальное качество достигается при 16 битах из которых 12 – значащие

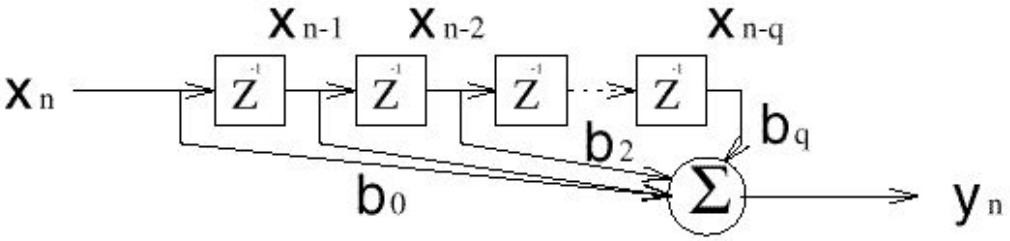
Линейные фильтры

Фильтры с конечной импульсной характеристикой

Фильтр с конечной импульсной характеристикой (КИХ) вычисляет выходное значение *y(n)*, как взвешенную сумму текущего входного значения и предыдущих входных значений.

$$Y_n = b_0 x_n + b_1 x_{n-1} + b_2 x_{n-2} + \ldots + b_q x_{n-q} = \sum_{j=0}^{q} b_j x_{n-j}$$

Блок-схема КИХ-фильтра



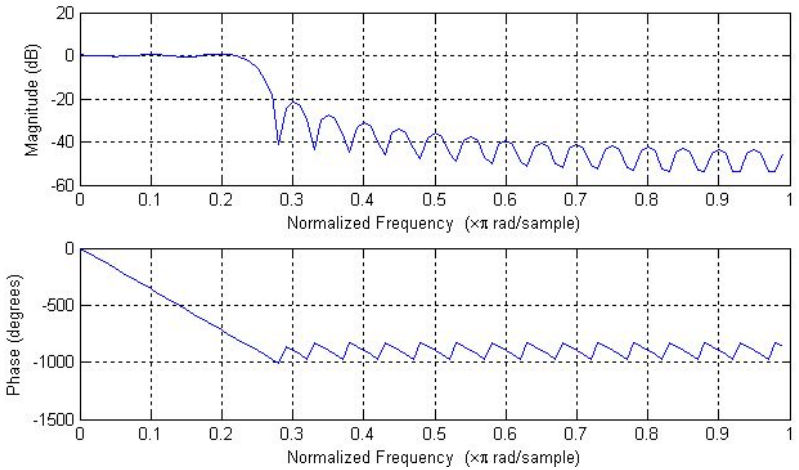
Передаточная характеристика КИХ-фильтра

$$H(w) = \sqrt{\left(\sum_{j=0}^{q} b_j \cos(-\omega j T)\right)^2 + \left(\sum_{j=0}^{q} b_j \sin(-\omega j T)\right)^2}$$

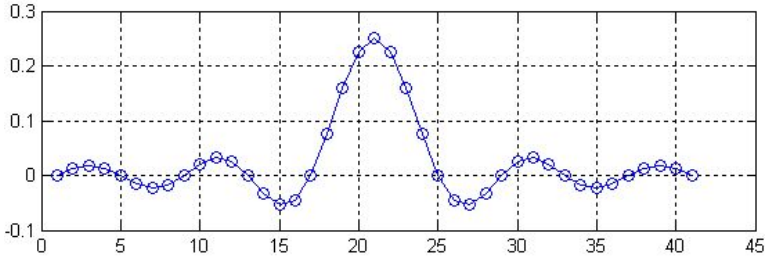
Линейные фильтры

Фильтры с конечной импульсной характеристикой

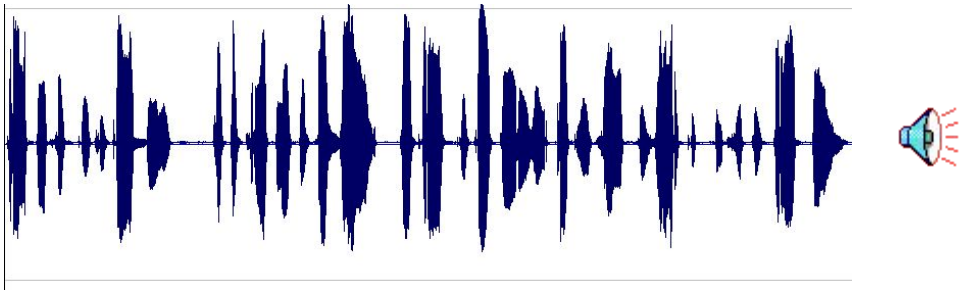
Амплитудно-частотная характеристика фильтра нижних частот



Импульсная характеристика фильтра нижних частот



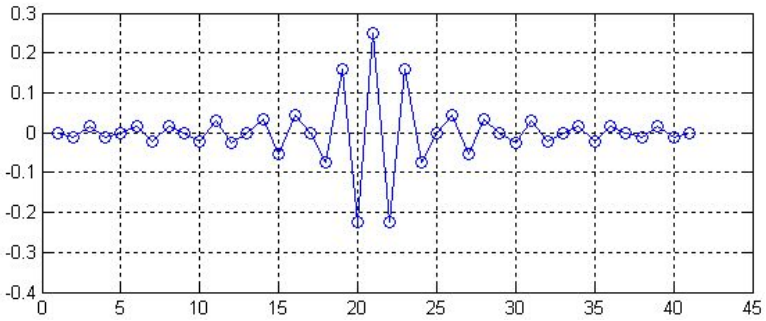
Сигнал после НЧ фильтрации



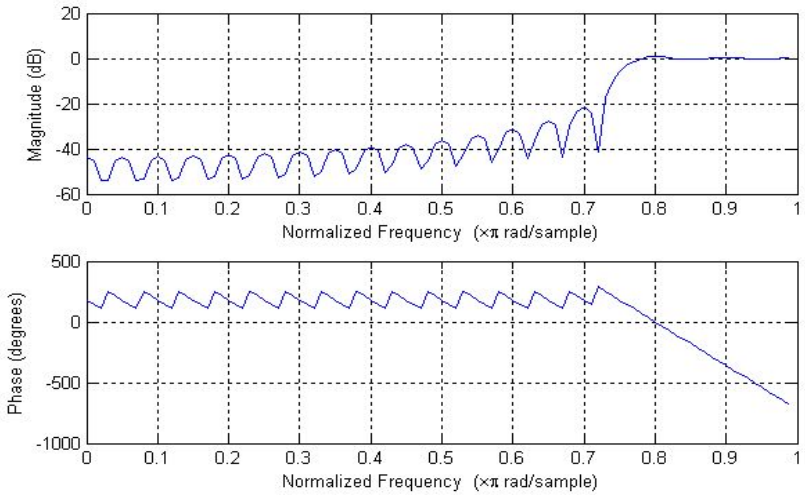
Линейные фильтры

Фильтры с конечной импульсной характеристикой

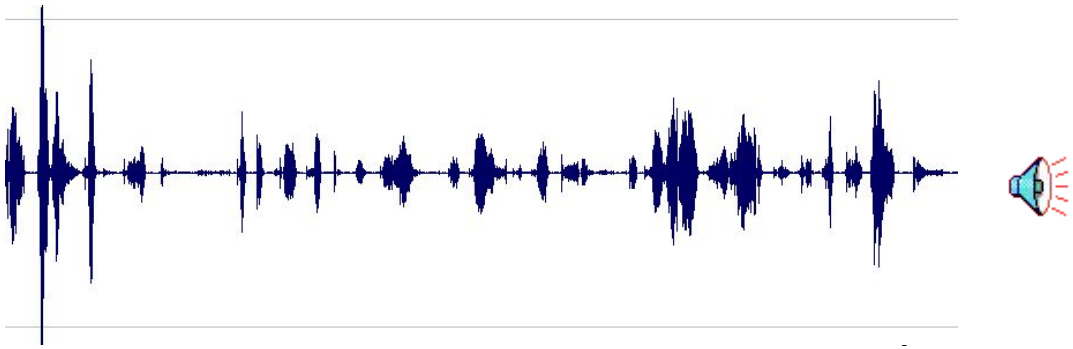
Импульсная характеристика фильтра высоких частот



Амплитудно-частотная характеристика фильтра высоких частот



Сигнал после ВЧ фильтрации



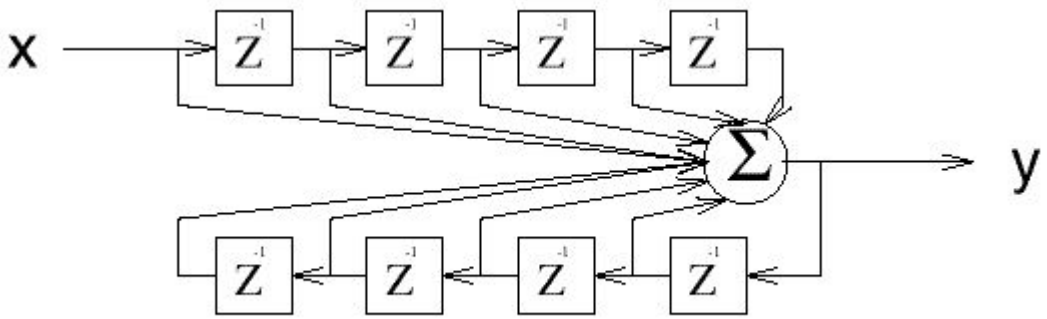
Линейные фильтры

Фильтры с бесконечной импульсной характеристикой

Фильтры с бесконечной импульсной характеристикой (БИХ) производят выходное воздействие, *y(n)*, как взвешенную сумму текущего и предыдущих входных воздействий, *x(n)*, и предыдущих выходных воздействий.

$$y_n = \sum_{i=0}^{p} a_i y_{n-i} - \sum_{j=0}^{q} b_i x_{n-j}$$

Блок-схема БИХ-фильтра



Обычные типы фильтров

Тип фильтра	Характеристика
Баттерворта	Максимально плоская амплитуда
Бесселя	Максимально плоская групповая задержка
Чебышева	Равноволнистый в области пропускания или области останова
Эллиптически	Равноволнистый в области пропускания и области останова

Линейные фильтры

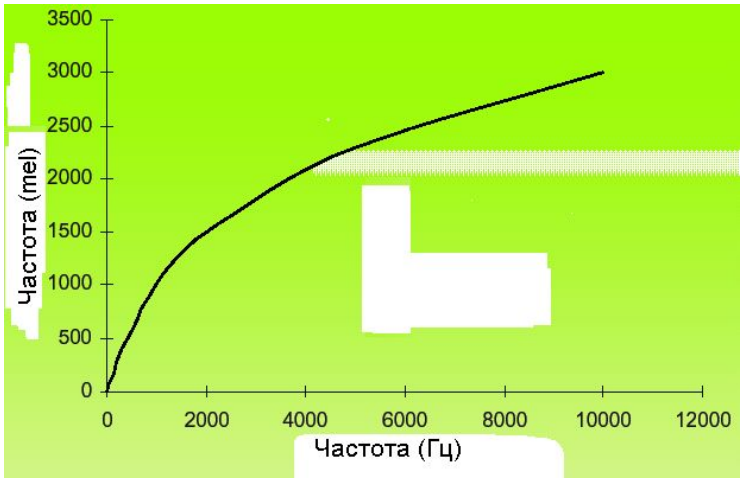
Анализ банка фильтров

Информативность различных частей линейного спектра неодинакова: в низкочастотной области содержится больше информации чем в высокочастотной. Поэтому для предотвращения излишнего расходования ресурсов, необходимо уменьшать число элементов, получающих информацию с высокочастотной области, или, что то же самое, сжать высокочастотную область спектра в пространстве частот. Наиболее распространенный метод – логарифмическое сжатие или приведение к mel шкале:

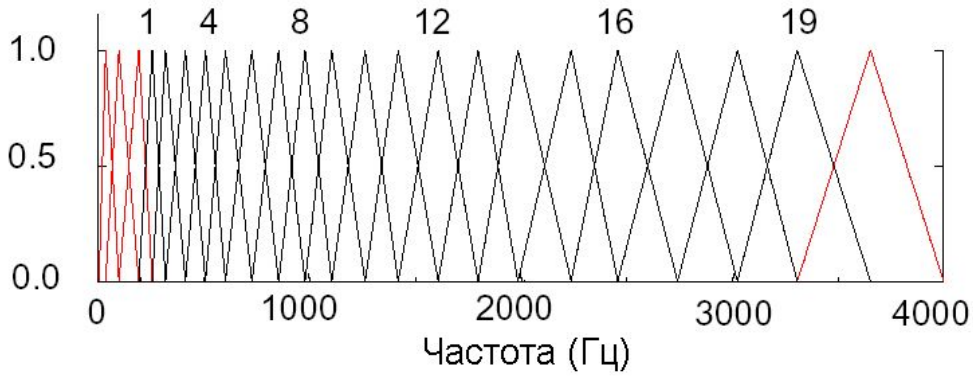
$$m = 1125 \log(0.0016 f + 1)$$

f - частота в спектре, Гц, а

m - частота в новом пространстве, mel



Mel-шкала



Банк фильтров

Кратковременный анализ Фурье

Дискретное преобразование Фурье (ДПФ)

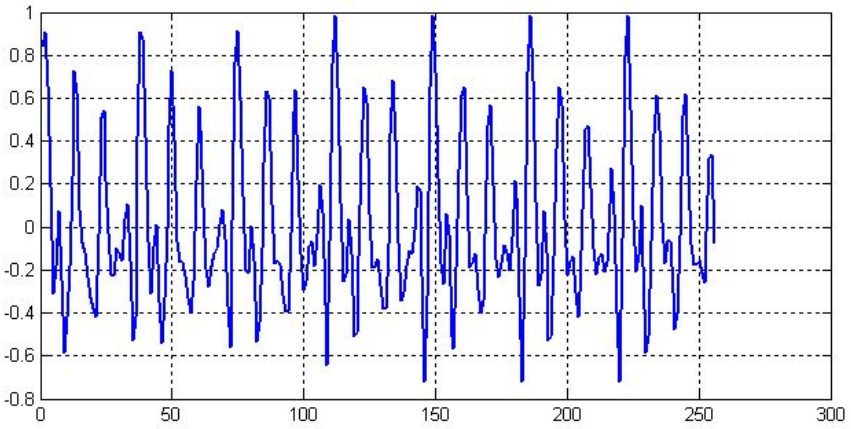
$$X(e^{i\theta}) = \sum_{n=-\infty}^{n=\infty} x(n)e^{-in\theta}$$

Где $\theta = 2\pi f T = 2\pi f / f_s$, T – период дискретизации, f_s – частота дискретизации.

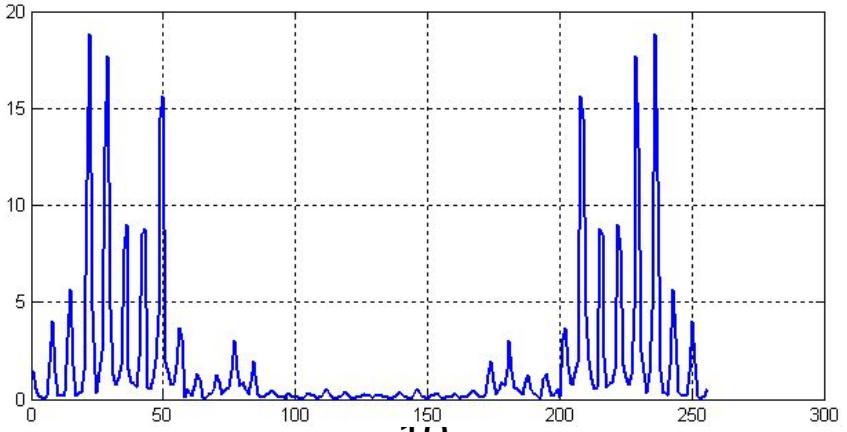
Обратное преобразование Фурье

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{i\theta}) e^{in\theta} d\theta$$

Сигнал звука «а» в t-области



Сигнал звука «а» в частотной области



Кратковременный анализ Фурье

Свойства ДПФ

- Линейность

$$\alpha h(n) + \beta g(n) \leftrightarrow \alpha H\left(e^{i\theta}\right) + \beta G\left(e^{i\theta}\right)$$

- Временной сдвиг

$$h(n+k) \leftrightarrow H\left(e^{i\theta}\right)e^{i\theta k}$$

- Частотный сдвиг

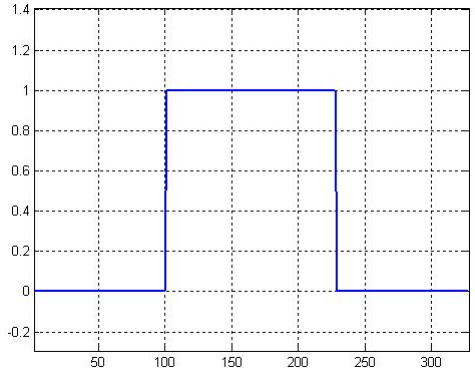
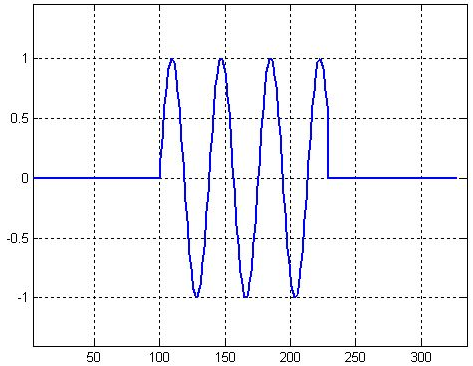
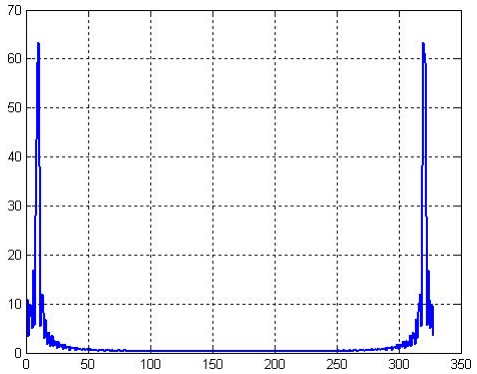
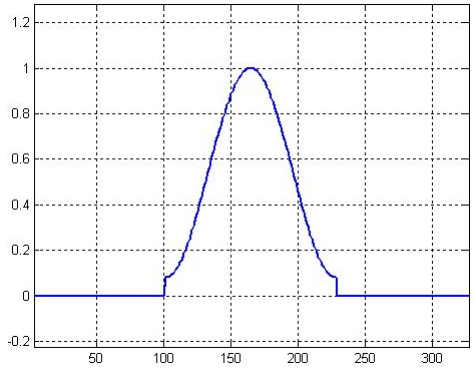
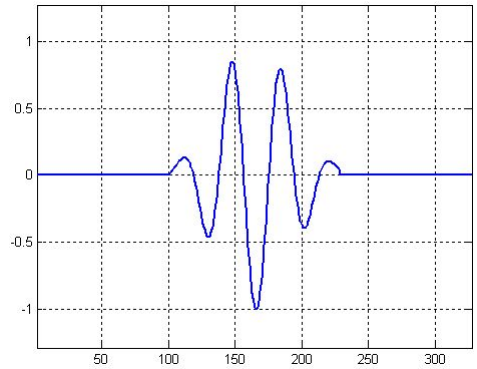
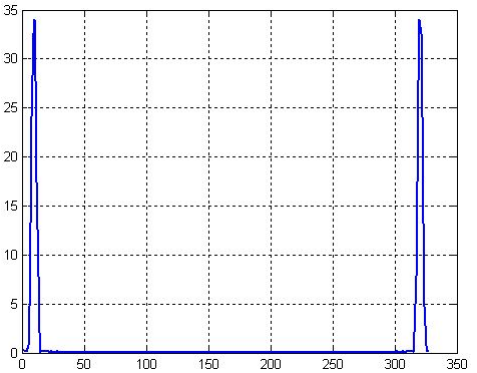
$$h(n)e^{in\theta_0} \leftrightarrow H\left(e^{i(\theta-\theta_0)}\right)$$

- Свертка

$$h(n) * g(n) \leftrightarrow H\left(e^{i\theta}\right)G\left(e^{i\theta}\right)$$

$$h(n)g(n) \leftrightarrow H\left(e^{i\theta}\right) * G\left(e^{i\theta}\right)$$

Применение окон

	Вид окна во временной области	Сигнал после наложения окна	Спектр сигнала
Прямоугольное окно			
Окно Хэмминга			

Умножение сигнала на функцию окна во временной области равносильно свертке сигнала в частотной области

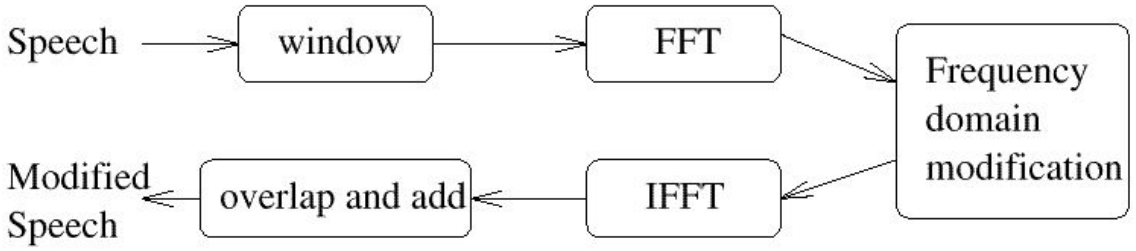
Применение окон

Наиболее часто используемые окна

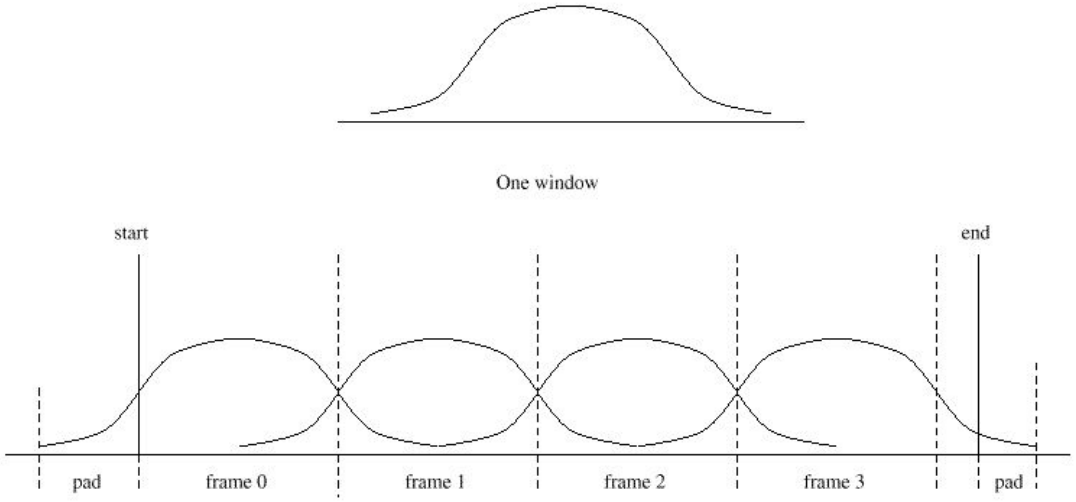
- Прямоугольное
- Треугольное
- Хэмминга
- Блэкмана
- Блэкмана-Харриса
- Ханна
- Чебышева
- Гаусса
- Кайзера

Применение окон

Метод перекрывания и добавления в линейной фильтрации



Данное окружение подходит в большинстве задач фильтрации, где фильтр может зависеть от времени и анализируемого сигнала



Метод перекрывания и добавления во временной области

Кодирование речи

- Речь может быть закодирована на многих уровнях
- Низкий Bit-rates достигается путем наложения больших ограничений на механизм получения речи.
- Качество уменьшается с уменьшением bit-rate

Waveform кодеры

Импульсная кодовая модуляция (PCM)

- Требуется, чтобы частота дискретизации, *fs*, была больше частоты Найквиста (в два раза большая, чем максимальная частота сигнала)

 wav-файл (106kbyte)

Дифференцированная импульсная кодовая модуляция (DPCM)

- Предсказывает следующий отсчет, основываясь на нескольких отсчетах, *декодированных* последними
- Минимизирует среднеквадратичную ошибку остатка предсказания – использует LP-кодирование.

Адаптивная дифференцированная импульсная кодовая модуляция (ADPCM)

- Адаптируется предсказатель
- Предшествующая адаптация: новые значения предсказания уточняются из входных данных
- Последующая адаптация: используются значения предсказателя, вычисленные из недавно декодированного сигнала

 vox-файл (26kbyte)

Кодирование речи

Кодировщики подобластей

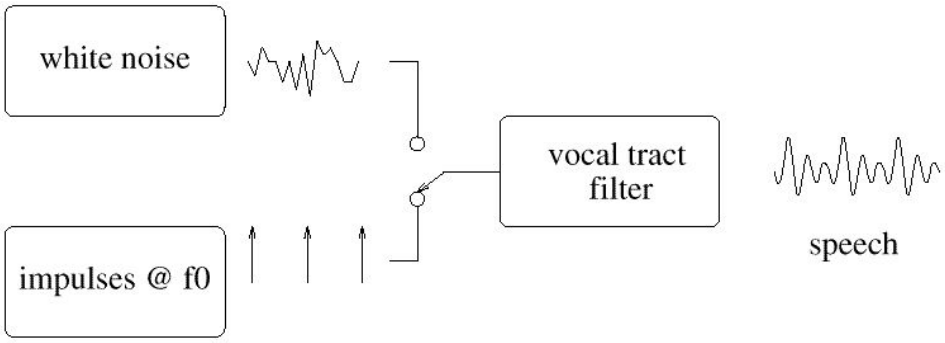
- Использует неравномерную частотную чувствительность слуховой системы.
- Каждая подобласть кодируется со свойственной ей разрешением – например 4 бита на отсчет в низкочастотной подобласти и 2 бита на отсчет в высокочастотной подобласти.
- Также может использоваться слуховое маскирование – используется меньше бит если соседняя подобласть намного громче.
- Основа для стандарта MPEG-audio (сжатие 5:1 с CD качеством звука без заметной деградации).

Пример: MP3 32, 64, 128, 256, 320 kbit/sec

Кодирование речи

Вокодеры линейного предсказания

- Для каждого фрейма необходимо закодировать:
  - Представление LP-фильтра
  - Мощность
  - Затухание голоса
  - Высоту (если есть голос)
- Большинство битов идет на LP-параметры
- Обычно используют «LP-коэффициенты» или «Линейные спектральные пары» для представления LP-параметров:



CELP кодеры

Кодеры, возбуждаемые кодами линейного предсказания (Code Excite Linear Prediction)

- Основан на базисном LP-кодере
- Применяется долговременный предсказатель для устранения избытка повторяемости
- Кодирование требует намного больших вычислительных затрат чем декодирование (нужен поиск в codebook).
- Результирующий bit-rate около 4 kbps.

G.729 (8 kbit/sec)

ICELP (4.8 kbit/sec)

MMBE (2.4 kbit/sec)

LBRAMR (1.2 kbit/sec)

Спасибо за внимание