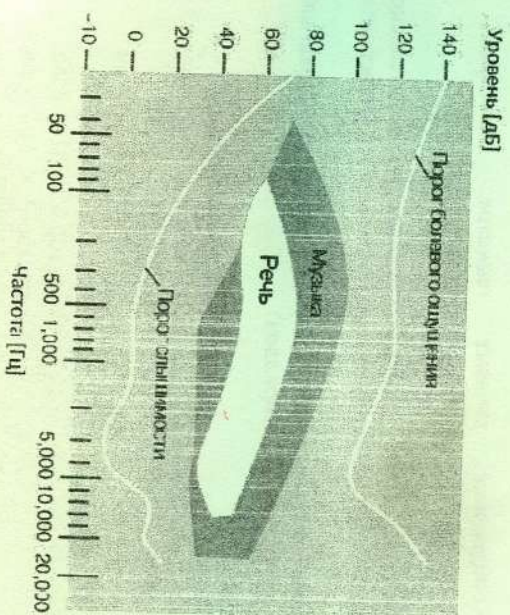


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. В.Ф. УТКИНА

КОДЕКИ РЕЧЕВЫХ СИГНАЛОВ

Методические указания
к лабораторным работам



УДК 621.395.01

Кодексы речевых сигналов: методические указания к лабораторным работам / Рязан. гос. радиотехн. ун-т им. В.Ф. Уткина; сост. С.Н. Кириллов, В.Т. Дмитриев. Рязань, 2020. 48 с.

Изложены основные элементы теории по кодированию сигналов по средствам вейвлет-преобразования, принципам работы вокодеров при кодировании речевых сигналов, оценке качества речевых сигналов в МТС и использовании вейвлет-преобразования для обработки речевых сигналов. Даны методические указания для подготовки и выполнения лабораторных работ.

Предназначены для бакалавров по направлению «11.03.02 Информационные технологии и системы связи», магистрантов по направлению «11.04.02 Информационные технологии и системы связи» и специалистов по специальности «11.05.01 Радиоэлектронные системы и комплексы».

Ил. 21. Библиогр.: 13 назв.

Кодирование речевых сигналов, вокодеры, полосовое кодирование, вейвлет-преобразование

Печатается по решению редакционно-издательского совета Рязанского государственного радиотехнического университета.

Рецензент: кафедра радиоуправления и связи (зав. кафедрой д-р техн. наук, проф. С.Н. Кириллов)

Кодексы речевых сигналов

Составители:
Кириллов Сергей Николаевич
Дмитриев Владимир Тимурович

Редактор М.Е. Цветкова

Корректор Р. К. Манутова

Подписано в печать 30.07.2020. Формат бумаги 60х84 1/16.
Бумага писчая. Печать трафаретная. Усл. печ. л. 3,0.

Тираж 50 экз. Заказ 3857
Рязанский государственный радиотехнический университет,
390005, Рязань, ул. Гагарина, 59/1.
Редакционно-издательский центр РГРТУ.

Лабораторная работа № 1 Изучение принципов работы вокодеров при кодировании речевых сигналов

Цель работы: изучение устройства и принципа функционирования современных вокодеров, дискретного преобразования Фурье и особенностей спектрального анализа речевого сигнала.

Теоретическая часть

Для повышения эффективности использования пропускной способности канала связи в цифровых системах передачи может использоваться **полосное кодирование**, которое заключается в разделении полосы частот речевого сигнала $s(t)$ на ряд небольших полос с помощью полосовых фильтров ПФ, и раздельном их кодировании в тракте передачи.

При этом небольшие по ширине полосы частот каналы называются **субканалами**, а их количество зависит от требуемой степени сжатия и качества речевого сигнала и обычно лежит в пределах от 4 до 8.

После полосового фильтра сигнал каждого из субканалов с помощью однополосной модуляции переносится в полосу частот $0 \dots F_{\text{max}}$, а потом кодируется с помощью обычного ИКМ или АДИКМ-кодека. Следует заметить, что частота дискретизации $F_{\text{д}}$, а также число уровней квантования для каждого субканала выбирается индивидуально с учётом психоакустических свойств данного участка спектра речевого сигнала.

Рассмотрим **основные причины**, по которым полосное кодирование обеспечивает лучшую степень сжатия по сравнению с другими методами кодирования речевого сигнала:

- в зависимости от ценности восприятия для человеческого уха разных частот в каждом из субканалов можно использовать различное число уровней квантования, а значит, и различную скорость передачи

(так, низкочастотная часть спектра должна передаваться с большей точностью, поскольку в данной части спектра речевого сигнала располагаются основные форманты);

- в каждом из субканалов шаг квантования выбирается индивидуально в зависимости от уровня сигнала в данном субканале: при большом уровне речевого сигнала на данной частоте шаг квантования может быть выбран также большим.

Устройства кодирования, учитывающие статистические свойства речевого сигнала, называются **вокодерами**. Как правило, большинство вокодеров используют особенности спектра речевого сигнала и делится на следующие виды:

- **полосные вокодеры**;
- **ортогональные (гармонические) вокодеры**;
- **формантные вокодеры**;

- **вокодеры с линейным предсказанием**.

Для речевого сигнала характерны следующие особенности спектра:

- спектр гласных и звонких согласных звуков является дискретным, причем все компоненты спектра являются кратными гармониками частоты основного тона F_0 , индивидуального для каждого человека.

$$s(t) = \sum_{k=1}^N A_k \cos(2\pi k F_0 t), \quad (1.1)$$

где A_k — амплитуда k -й гармоники основного тона F_0 ;

- спектр глухих согласных звуков является практически сплошным;

- для всех звуков характерно неравномерное распределение энергии спектральных составляющих с концентрацией их на отдельных частотах, называемых формантами. Число формант для каждого звука зависит от особенностей голосового аппарата человека и лежит в пределах от 3 до 5, а иногда доходит даже до 7. Основными параметрами каждой форманты являются ее центральная частота F_{ϕ} и максимальная

амплитуда A_{ϕ} , которые изменяются от звука к звуку и от говорящего к говорящему;

- усредненные центральные частоты первых трех формант лежат в пределах: первая — 200...700 Гц, вторая — 1000...2000 Гц и третья — 2000...4000 Гц;

- частота следования отдельных звуков не превышает 10 Гц, причем длительность самых коротких согласных звуков составляет не менее 30 мс, а самых длинных гласных — не более 350 мс;

- фазовые соотношения между отдельными частотами спектра речевого сигнала менее существенны для восприятия, поскольку ухо воспринимает не столько фазовые различия между отдельными частотными составляющими, сколько уровень энергии отдельных частот.

Полученный аналоговый сигнал $s_{\text{ан}}(t)$ разделяется на ряд субканалов с помощью полосовых фильтров ПФ. На выходе каждого полосового фильтра установлен амплитудный детектор АД, основной задачей которого совместно с ФНЧ (с полосой пропускания от 0 до 25...35 Гц) является выделение огибающей сигнала данного субканала.

Дело в том, что сигнал огибающей имеет достаточно низкую частоту и содержит информацию о средней интенсивности спектра в пределах полосы частот данного субканала. Далее сигнал огибающей спектра подвергается кодированию с использованием классической ИКМ или АДИКМ и поступает на устройство объединения УО. Блок **выделителя основного тона (ВОТ)** предназначен для определения частоты основного тона F_0 речевого сигнала. **Определитель тон-шум (ОТШ)** используется для определения типа звука — звонкий (тональный) или глухой (шумоподобный). Для передачи всей информации от блоков ВОТ и ОТШ требуется полоса частот порядка 100 Гц. Блоки ВОТ, ОТШ и оборудование субканалов совместно образуют **анализатор речевого сигнала**.

Всего для передачи сигналов ВОТ, ОТШ и сигнала цикловой синхронизации требуется порядка 600 бит, а при числе субканалов от 6 до 12 скорость передачи полосного вокодера составляет 1200...2400 кбит/с. На приемной стороне декодерный групповой

цифровой сигнал (ВПС) поступает на устройство разделения, после чего он разделяется на сигналы отдельных субканалов, причем информация от блока ВОТ тракта передачи поступает на импульсный генератор основного тона (ГОТ) для управления его частотой, а информация от блока ОТП тракта передачи поступает на коммутатор К, задача которого заключается в подключении ко входам полосовых фильтров субканалов (амплитудным модуляторам АМ) либо ГОТ, либо генератора шума (ГШ). Генератор шума ГШ формирует короткие импульсы случайной длительности и частоты следования, что эквивалентно шуму с равномерной спектральной плотностью.

Генератор основного тона представляет собой генератор гармоник, формирующий периодическую последовательность импульсов, следующих с частотой основного тона $F_{от}$. Совокупность всех перечисленных выше блоков приемного тракта называется **синтезатором речевого сигнала**.

Отбрасывая сигнала соответствующего канала с выхода ЦАП поступает на **амплитудный модулятор АМ**, где происходит амплитудная модуляция либо сигналов ГОТ, либо сигналов ГШ. В результате получается так называемый окрашенный спектр частот, характерный для того или иного звука. Иными словами, в зависимости от сигнала отбрасываемой отдельного субканала меняется интенсивность спектра в его полосе, а полосовые фильтры используются в основном для устранения побочных продуктов модуляции. После указанных преобразований сигналы с выходов полосовых фильтров объединяются и поступают на усилитель УС, обеспечивающий номинальный уровень сигнала на приеме.

Основной идеей гармонического вокодера является передача отбрасываемой текущего энергетического спектра речевого сигнала, усредненного за некоторый, обычно небольшой, промежуток времени, в виде коэффициентов разложения отбрасываемой в ряд Фурье. В зависимости от требуемой точности воспроизведения отбрасываемой, а значит, и требуемого качества речевого сигнала количество членов ряда будет различным. Поскольку каждый член ряда определяется как минимум двумя параметрами, то число параметров будет всегда в два

раза больше числа членов ряда. Так, если число членов ряда Фурье будет равно n , то соответственно число параметров, которые требуются передать, составит $2n-1$.

На практике задача сводится к следующему: требуется измерять уровень спектра речевого сигнала на $(2n-1)$ частотах, выделяемых полосовыми фильтрами. Величины этих уровней и будут представлять собой параметры передачи. Оптимальная ширина полосы пропускания полосового фильтра может быть вычислена по следующей формуле:

$$\Delta f \leq \frac{\Delta f}{2q}, \quad (1.2)$$

где Δf — полоса частот речевого сигнала (обычно выбирается равной $0,3 \dots 3,4$ кГц); q — число формант.

Спектральные составляющие речевого сигнала изменяются достаточно медленно — примерно с частотой $25 \dots 50$ Гц. Поскольку для сохранения приемлемого качества речи требуется передача информации не менее чем о $3 \dots 4$ формантах, то число значений спектральных составляющих должно быть не менее $6 \dots 8$. Для этого потребуются $3 \dots 5$ членов разложения в ряд Фурье, что потребует полосы частот гармонического вокодера $75 \dots 250$ Гц.

Формантный вокодер определяет амплитуду и расположение частотных максимумов речевого сигнала — формант и передает эту информацию вместе с отбрасываемой всего спектра речевого сигнала. Таким образом, формантный вокодер концентрирует внимание на передаче лишь наиболее значимых составляющих спектра и отличается от полосного вокодера построением анализатора и синтезатора, хотя блоки выделения основного тона и сигнала тон-шум практически одинаковы, но число субканалов уменьшается до числа формант (обычно не более трех).

При этом главным требованием к формантному вокодеру является точное отслеживание динамики изменения формант во времени. Если данное требование выполняется, то приемлемое качество речи достигнимо уже на скоростях не более 1000 бит/с.

Вокодеры с линейным предсказанием в отличие от рассмотренных ранее полосного, гармонического и формантного

вокодеров анализируют сигнал не в частотной (спектр), а во временной области. Основной целью анализа речевого сигнала является получение меняющейся во времени модели возбуждения ГОТ и ГШ синтезатора тракта приема и функции передачи. При этом синтезатор тракта приема формирует речевой сигнал путем пропуска сигнала полученных сигналов ГОТ и ГШ через устройство, соответствующее математической модели речеобразующего тракта (рис. 1.1).

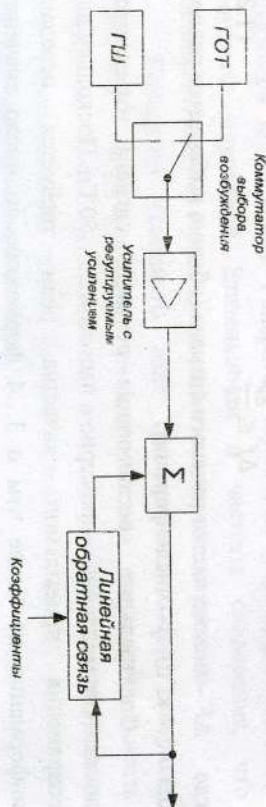


Рис. 1.1. Структурная схема вокодера с линейным предсказанием

Синтезатор периодически приспосабливается к изменением параметров модели и элементов возбуждения путем постоянного уточнения этих параметров, но в любой момент времени речевой тракт представляет собой линейное устройство с постоянными параметрами в течение данного интервала уточнения:

$$y(n) = \sum_{k=1}^p A_k y(n-k) + Kx(n), \quad (1.3)$$

где $y(n)$ – n -й отсчет выходного сигнала; A_k – k -й коэффициент предсказания; K – коэффициент усиления регулируемого усилителя УС; $x(n)$ – входной сигнал в момент дискретизации n ; p – порядок модели, под которым понимается количество последних отсчетов, используемых для линейного предсказания текущего отсчета.

Таким образом, выходной сигнал в момент времени n представляет собой сумму входного сигнала в данный момент и линейной комбинации p предыдущих отсчетов выходного сигнала. Адаптивность данной модели заключается в том, что кодер

периодически выбирает оптимальные коэффициенты предсказания A_k для достижения минимума среднеквадратического значения ошибки.

На передающей стороне вокодера с линейным предсказанием в анализаторе определяются и передаются следующие **параметры речевого сигнала**:

- информация о характере возбуждения (тон или шум);
- период основного тона (для воспроизведения гласных и звонких согласных);
- коэффициент усиления K ;
- коэффициенты предсказания A_k (параметры модели речеобразующего тракта).

Для определения оптимальных коэффициентов предсказания используется следующая методика: в качестве желаемого выходного сигнала синтезатора выбирается реальный выходной сигнал, затем составляется система p линейных уравнений с p неизвестными. Решение данной системы уравнений дает значения искоемых коэффициентов предсказания. При скорости передачи 1,2...2,4 кбит/с вокодеры с линейным предсказанием обеспечивают приемлемое качество речи, даже более естественное звучание, чем другие вокодеры. Рассмотренные выше вокодеры, несмотря на высокую степень сжатия, не обеспечивают желаемый уровень разборчивости и натуральности звучания речи. Причиной характерного машинного звучания, широко используемого в музыкальной индустрии, является неточное определение частоты основного тона, отпавляющей спектра на низких частотах. А особенностью человеческого уха является то, что оно способно подмечать мельчайшие отличия именно основного тона речевого сигнала.

В полувокодере информация об основном тоне передается непосредственной полосой низкочастотной части спектра речевого сигнала, образуя нулевой субканал на выходе ФНЧ. Сигналы нулевого субканала и остальных субканалов анализатора объединяются и передаются на приемную сторону в виде группового цифрового сигнала.

На приеме происходят обратные преобразования: сначала разделение сигналов субканалов в устройстве распределения УР, затем цифровальное преобразование, а после этого сигнал нулевого субканала подается на устройство формирования широкополосного сигнала (ФШС) с равномерным спекром. Данный сигнал с равномерным спекром используется как источник возбуждения синтезатора, формирующего спектр речевого сигнала выше нулевого субканала (рис. 1.2).

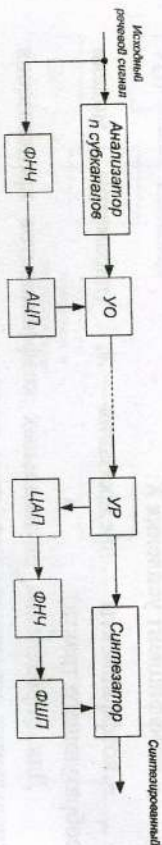


Рис. 1.2. Структурная схема полукодера

Детали временной структуры речевого сигнала на низких частотах при этом не теряются, а в точности воспроизводятся на приемной стороне. А поскольку именно они определяют натуральность речи, то восстановленный сигнал получается очень качественным, даже сравнимым по качеству с каналом тональной частоты 0,3...3,4 кГц.

Достижения современной микропроцессорной техники и микроэлектроники, а также алгоритмов цифровой обработки сигналов позволяют создавать компактные вокодеры со скоростями передачи 16, 9,6, 8 и 4,8 кбит/с, что в конечном счете позволяет еще больше увеличить эффективность использования пропускной способности линий связи.

Для выполнения лабораторной работы требуется выделить формант из частотного спектра речевого сигнала. Для выполнения данной задачи предлагается использовать дискретное преобразование Фурье.

Дискретное преобразование Фурье (ДПФ или DFT) вычисляется по следующей формуле:

$$S(k) = \sum_{n=0}^{N-1} s(n) e^{-j \frac{2\pi}{N} nk}, \quad (1.4)$$

где k – номер отсчета спектра сигнала, принимает значения от 0 до $N-1$.

Обратное дискретное преобразование Фурье (ОДПФ или IDFT) вычисляется по следующей формуле:

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{j \frac{2\pi}{N} nk}, \quad (1.5)$$

где n – номер отсчета сигнала во временной области, принимает значения от 0 до $N-1$.

Порядок выполнения работы

1. Запустить среду программирования Matlab R2017a.
2. Считать данные из входного аудиофайла, содержащего речевые данные, с помощью следующего блока программы:


```
voice=audioread('a.wav');
[y, Fs]=audioread('a.wav');
[q, n]=size(voice);
```

3. Задать разрядность квантования, размер окна, а также количество формант с помощью следующего блока программы:

```
N=512;
m=2;
formant=69;
```

4. С помощью следующего блока программы преобразовать последовательность отсчетов файла в двумерный массив, в котором исходная последовательность разбита на блоки по N отсчетам:

```
%Rvoice=zeros(N,q);
for j=1:q
    lvoice(1:N,j)=voice(1+N*(j-1):j*N,1);
    %Rvoice(1:N,j)=voice(1+N*(j-1):j*N,2);
end
```

5. Выполнить дискретное преобразование Фурье (ДПФ) для исходного сигнала, разбитого на блоки по N отсчетов каждый. После этого выполнить квантование с помощью следующего блока программы:

9. Сделать обратное преобразование Фурье с помощью следующего блока программы:

```
for j=1:q
    cutldst(1:N,j)=cutldst(1:N,j)*(max(cutldst(1:N,j))-1e-15)/2^m;
    %cutldst(1:N,j)=cutldst(1:N,j)*(max(cutldst(1:N,j))-1e32)/2^m;
end
ldstnew=cutldst.*exp(1i*rldst);
%Rldstnew=cutldst.*exp(1i*rldst);
lvosvoice=ifft(ldstnew);
%Rvosvoice=ifft(Rldstnew);
vosvoice=zeros(q,n);
for j=1:q
    vosvoice(1+N*(j-1):j*N,1)=round(lvosvoice(1:N,j));
    %vosvoice(1+N*(j-1):j*N,2)=round(Rvosvoice(1:N,j));
end
```

10. Построить фильтр выходного сигнала с помощью следующего блока программы:

```
w = 3;
Wn = (100/3000)/(Fs);
[b,a] = butter('w',Wn);
vosvoice = filter(b,a,vosvoice);
```

11. Сохранить полученный сигнал в новый файл с помощью следующей функции:

```
audiowrite('vosvoice.wav',vosvoice,Fs);
```

12. Прослушать полученный аудиофайл.

Примечание: в случае использования файла с двумя звуковыми дорожками (стерео) нужно удалить символ «%» в начале каждой строки, тем самым задействовать строки в работе программы.

Контрольные вопросы

1. Полосное кодирование речевого сигнала.
2. Статистические особенности речевого сигнала.
3. Устройство и принципы функционирования полосного вокодера.
4. Устройство и принципы функционирования гармонического вокодера.

5. Устройство и принципы функционирования формантного вокодера.

6. Устройство и принципы функционирования вокодера с линейным предсказанием.

7. Особенности современных вокодеров, используемых в системах связи. Вокодеры *CELL*, *ASCELL*.

8. Способы улучшения качества звучания вокодеров. Полувокодеры.

9. Дискретное преобразование Фурье. Быстрое преобразование Фурье.

Содержание отчета

1. Цель работы.
2. Структурная схема разрабатываемого и исследуемого вокодера.
3. Изображение графиков, построенных в ходе выполнения работы.
4. Выбор значений полосы пропускания выходного фильтра и обоснование выбора.
5. Выводы по работе.

Библиографический список.

1. Куанышев А. Н. Речевое кодирование // Colloquium-journal. 2020. № 7. 59 с.
2. Подвальный С.Л., Рошупкин А.Д. Обзор методов и алгоритмов сжатия речевой информации в системах цифровой радиосвязи // Вестник ВГТУ. 2017. № 2. С. 7-13.
3. Носов Л.С., Вечерский В.В., Зудин В.С., Можайкин А.В. Кодирование речевой информации в системах IP-телефонии // Вестник Сыктывкарского университета. Серия 1. Математика. Механика. Информатика. 2016. № 21. 86 с.

Лабораторная работа № 2

Оценка качества речевых сигналов в МТС

Цель работы: изучение принципов оценки качества речевых сигналов в МТС, изучение влияния разрядности квантования, размера окна и количества формант на качество речевых сигналов.

Теоретическая часть

В общем случае под речевым сигналом в телекоммуникационных системах подразумевается процесс, выполняющий функцию передачи речевого сообщения. Это могут быть акустические, механические, электрические и другие процессы. Речевое сообщение обычно создается в мозгу человека. Затем оно превращается в сигналы нервной системы, управляющие артикуляционными движениями органа речи. В свою очередь, эти движения управляют формированием речевых сигналов в речеобразующем тракте. Речеобразующий тракт состоит из ларингальных, фарингальных, ротовых и носовых полостей, объема и упругость стенок которых меняются во времени. Изменения конфигурации речеобразующего тракта в процессе произнесения звуков речи воздействуют на проходящую через тракт акустическую волну. Полученные сигналы излучаются в окружающее пространство в виде акустического сигнала. Далее акустический сигнал в тракте передачи телекоммуникационной системы преобразуется в электрический сигнал.

В дальнейшем сигнал может подвергаться различным преобразованиям, оставаясь электрическим. В одних случаях электрический сигнал после усиления непосредственно поступает в электроакустический преобразователь, в других случаях после ряда преобразований по форме (модуляции и демодуляции, квантование, кодирование, компандирование и т. п.) в конце концов тоже, как правило, превращается в акустический сигнал, однако сигнал может остаться цифровым, например, при задачах распознавания речи. Акустический сигнал, воздействуя на барабанную перепонку уха, превращается в механический сигнал, а во внутреннем ухе — в сигнал

нервной системы. Этот сигнал в центральной нервной системе как бы расшифровывается, в результате чего воссоздается первоначальное сообщение. В ряде случаев это сообщение может отличаться от исходного, что связано с искажениями в системах связи.

Частота звуковой волны может варьироваться очень сильно, но человек воспринимает частоты от 20 до 22 000 Гц (длина волны от 1,56 см до 17,19 м). Динамический диапазон уровней громкости, воспринимаемых человеком, очень большой.

Экспериментально установлено, что мощность самого слабого различимого звука составляет 1 мкВт. Существует также шкала громкости в единицах звукового давления, где за нулевой уровень принято давление 10^{-5} Н/см². Для этой шкалы используется обозначение dB SPL (Sound Pressure Level - SPL).

Практически передача речевого сообщения по электрической связи решается посредством цифрового представления, обработки и передачи по каналу связи речевого сигнала — акустической волны, представленной в виде электрического колебания. В таком случае речь представляет собой колебания сложной формы, зависящей от произносимых слов, тембра голоса, интонации, пола и возраста говорящего. Под частотным диапазоном речи понимают ширину полосы частот, занимаемой речевым сигналом. Он заключается в диапазоне частот от 100 до 8000 Гц. Однако в соответствии с рекомендациями МККТТ (Международный консультационный комитет по телефонии и телеграфии) при преобразовании и обработке сигнала ограничиваются интервалом частот от 300 до 3400 Гц.

Субъективное восприятие звука. Как уже было отмечено, слуховой аппарат человека воспринимает звуки, частота которых находится в интервале от 20 до 22 000 Гц. Однако чувствительность человеческого уха не является одинаковой во всем воспринимаемом диапазоне. На рис. 2.1 приведены области слухового восприятия, в том числе соответствующие речи и музыке.

Частотный диапазон человеческой речи располагается приблизительно в интервале от 100 до 8000 Гц. Существование порога

слышимости является основой для построения алгоритмов сжатия звука с потерями.

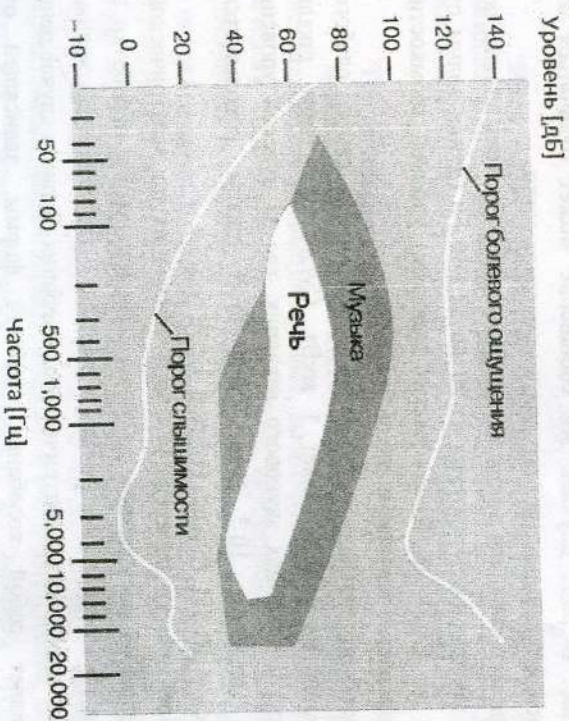


Рис. 2.1. Области слухового восприятия

Частотное маскирование. Кроме того, для эффективного сжатия используются ещё два свойства органов слуха человека: частотное маскирование и временное маскирование. Частотное (слуховое) маскирование происходит тогда, когда нормально слышимый звук накрывается другим громким звуком с близкой частотой. На рис. 2.2 схематично изображены маскирующий и маскируемый звуки.

Слышимый (маскирующий) звук приподнимает порог слышимости в своей окрестности (порог слышимости при маскировании). В результате звук, показанный пунктирной линией, становится не слышен, так как он маскируется более громким звуком. Это свойство используется при сжатии. Сигналы, соответствующие таким звукам, просто удаляются из массива данных, поскольку они все равно не будут услышаны. Частотное маскирование зависит от частоты

сигнала. Оно варьируется от 100 Гц на низких слышимых частотах до 4000 Гц на высоких. Следовательно, область слышимых частот может быть разделена на несколько критических полос, в пределах которых падает чувствительность уха.

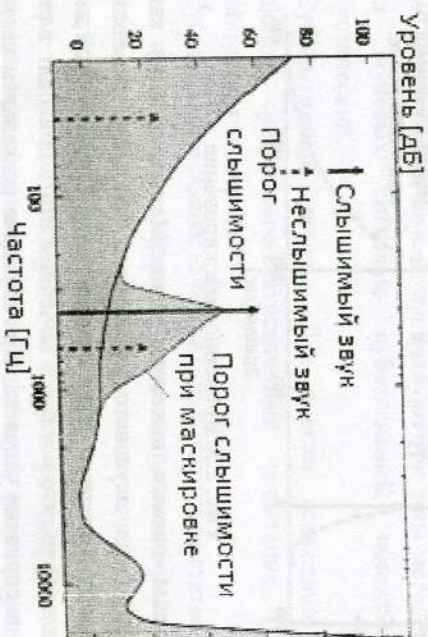


Рис. 2.2. Маскирующий и маскируемый звуки

Критические полосы. Критические полосы — ещё одна характеристика звука наряду с частотой. В отличие от частоты критические полосы определяются в соответствии со слуховым восприятием.

Критические полосы можно описать следующим образом: из-за ограниченности слухового восприятия звуковых частот порог слышимости частоты f приподнимается соседним звуком, если звук находится в критической полосе f . Ширина критической полосы называется её размером.

Временное маскирование. Временное маскирование наблюдается, когда громкому звуку частоты f по времени предшествует или за ним следует менее громкий звук близкой частоты (а также при одновременном звуке близкой частоты). Маскировка предшествующего звука проявляется на интервале не более 10 мс, тогда как последующий звук может быть маскирован на интервале от 100 до 200 мс (рис. 2.3).

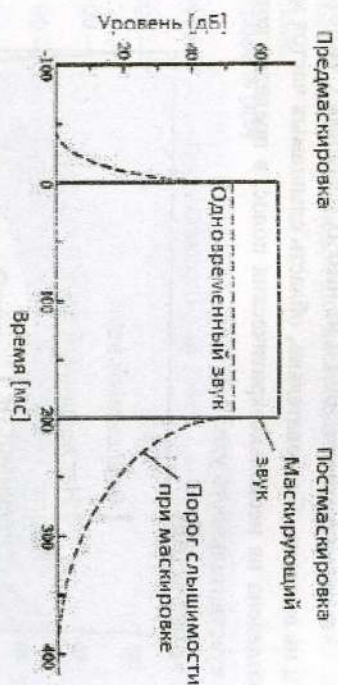


Рис. 2.3. Временное маскирование

Методы оценки качества речи. Методы оценки качества системы передачи звукового сигнала в основном определяются назначением этой системы (речь, музыка и т.п.). Общей чертой для этих систем является то, что все они в конечном итоге приходят к системе слухового восприятия человека. При передаче по каналам связи речь рассматривается как случайный процесс, характеристики которого определяются параметры передаваемого сигнала (динамический диапазон, полосу, отношение сигнал/шум). Все эти параметры поддаются измерению и могут быть точно определены. Однако, учитывая, что, с другой стороны, речевой сигнал воспринимается человеком, следует отметить, что с точки зрения человека речевой сигнал оценивается субъективно, подчиняясь законам психофизиологии слуха. Таким образом, мы получаем следующую картину: качество речевого сигнала объективно оценивается рядом показателей, которые никогда не смогут полноценно заменить систему восприятия звука человеком. Не учитывая нюансов непосредственно передачи речевого сигнала, можно сказать, что объективная оценка качества речи должна коррелироваться с субъективной оценкой.

Качество кодека может оцениваться двумя критериями:

- 1) сохранение объективного качества. Определение искажений с помощью аппаратных средств;
- 2) сохранение субъективного качества. Определение заметности искажений путём субъективно-статистических испытаний.

Поскольку строгие математические соотношения, устанавливающие связь между объективными параметрами звукового тракта и восприятием, до сих пор не получены, то ни один из существующих методов не может дать точной оценки качества звучания. В настоящий момент существуют три типа методов оценки качества звуковых сигналов: субъективный, объективный и психоакустический.

Субъективные методы. Для оценки искажений сигнала производится субъективно-статистические испытания (ССИ) с помощью группы экспертов.

Субъективные оценки качества основываются на статистической обработке субъективных оценок качества достаточно большого числа слушателей-экспертов. Эти оценки существенно зависят от возраста и пола диктора, скорости произнесения фраз и других обстоятельств. Тесты при получении субъективных оценок проводят с имитацией реальных условий, например посторонний шум, фоновая речь других людей и т. п. Количественные результаты этих тестов отображают усредненное качество, уровень усилий слушателя, разборчивость, естественность звучания.

Средняя экспертная оценка MOS (Mean Opinion Score) — это субъективное измерение, определенное в рекомендации R.800 МСЭ-T для оценки качества передачи в телефонных сетях.

Для различных задач могут использоваться разные 5-бальные шкалы оценок категорий.

Вычисленная по оценкам величина (средняя оценка мнений) обозначается символами MOS. Оценки MOS приведены в таблице.

Оценка MOS

Субъективная оценка качества звучания речи	Уровень восприятия речевой информации	Оценка
Отлично	Речь воспринимается полностью и без усилий	5
Хорошо	Речь воспринимается свободно, без ощутимых усилий	4
Удовлетворительно	Речь воспринимается с умеренными усилиями, наличие дефектов неоспорно	3
Плохо	Речь воспринимается с трудом	2
Очень плохо	Речь не воспринимается полностью или частично	1

Абсолютные значения MOS зависят от контекста тестов, на них оказывалось влияние различия в уровне знания языка и т. д.

Объективные методы. Объективные методы предлагают менее трудоемкие способы по сравнению с субъективными. Объективные методы основываются на оценке степени отличия кодированного и исходного сигналов.

Параметры многообразны и для различных типов кодирования могут отличаться принципиально. Несмотря на то, что некоторая корреляция между объективными показателями и субъективным качеством есть, однозначно судить о субъективном качестве нельзя, чем, собственно, и объясняется разный для каждого из кодексов способ оценки качества. Таким образом, очевидно, что универсального объективного метода оценки субъективного качества на данный момент не существует.

Соотношение сигнал/шум (SNR). Наиболее распространенной оценкой является соотношение сигнал/шум. Этот метод также называют критерием общего отношения сигнал/шум. Он учитывает общее отношение мощности сигнала и шума на всей длительности сигнала. Однако при низкой интенсивности полезного сигнала на каком-либо отрезке он может быть замаскирован другим участком сигнала с большей интенсивностью полезного сигнала, что в конечном итоге искажает оценку.

Сегментное соотношение сигнал/шум (segSNR). Является развитием метода соотношения сигнал/шум. В этом случае оценка

отношения сигнал/шум производится на интервалах от 15 до 20 мс, что позволяет получить более точную оценку в целом за счет того, что неравномерная интенсивность сигнала не исказит всей картины в целом.

Психоакустические методы. Психоакустические методы оценки качества учитывают особенности человеческого восприятия звука в целом и речи в частности. Особенность данных методов в том, что оценивается только субъективное качество сигнала с помощью технических и программных средств. Таким образом, строго говоря, они относятся к объективным методам, но построены с учетом особенностей субъективного восприятия звука человеком.

По мере развития способов устранения избыточности звуковых сигналов оценивать качество объективными параметрами становилось невозможным, а субъективно-статистические испытания оставались (и остаются) весьма трудоемкими. Устранение избыточности основывалось на особенностях восприятия звука и речи человеком, что послужило поводом для применения знаний об этих особенностях уже в методах оценки качества.

Задача любого метода оценки качества речевого сигнала в том, чтобы достичь высокой степени корреляции с субъективно-статистическими испытаниями, которая до сих пор остается наиболее точной оценкой качества речи.

Большинство методов основано на сравнении оригинального и кодированного сигналов с помощью некоторой психоакустической модели. Оценивается степень заметности искажений в кодированном сигнале для человека. Психоакустическая модель — это модель, которая преобразует звуковой сигнал в его внутреннее представление с точки зрения слухового аппарата человека, которое и сравнивается с внутренним представлением исходного сигнала. Наиболее распространенными являются взвешенное искажение спектра (WSS) и оценка PESQ, определенная в рекомендации МСЭ-Т Р. 862.

Оценка PESQ. Данный алгоритм представляет собой объективную методику определения качества речевой связи в телефонных системах, которая прогнозирует результаты субъективной

оценки качества этого типа связи слушателями-экспертами. Для определения качества передачи речи в PESQ предусмотрено сравнение входного, или эталонного, сигнала с его искаженной версией на выходе системы связи. Этот процесс схематично показан на рис. 2.4.

Результатом сравнения входного и выходного сигналов является оценка качества связи, которая аналогична усредненной субъективной оценке MOS (Mean Opinion Score), определяемой группой слушателей-экспертов согласно спецификации МСЭ-Т Р.800.

Оценки PESQ калибруются с использованием огромной базы данных оценок MOS.

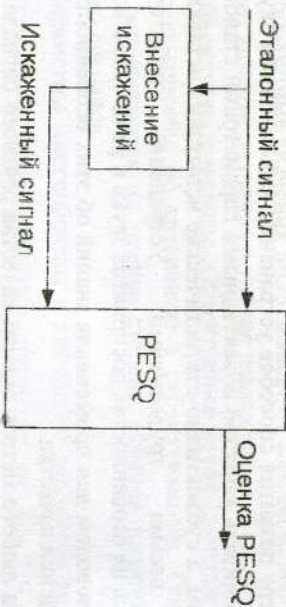


Рис. 2.4. Оценка PESQ

В PESQ заложено множество новых разработок, что выгодно отличает его от предыдущих алгоритмов оценки качества работы речевых кодеков, например PSQM и MNB [МСЭ-Т Р.861]. Эти инновации дают возможность уверенно использовать PESQ как для определения качества сквозной (end-to-end) передачи речи, так и для оценки влияния на качество связи отдельных элементов сетевого оборудования, включая кодеки.

На первом этапе PESQ вычисляется ряд задержек между первоначальным входом и ухудшенным выходом для каждого интервала, в котором задержка отличается от предыдущего временного интервала. Для каждого из этих интервалов рассчитываются соответствующие точки начала и окончания. Алгоритм юстировки (выравнивания) основан на правиле сравнения вероятности наличия

двух задержек в некотором временном интервале с вероятностью наличия единственной задержки для этого интервала. Алгоритм может обрабатывать задержки и в течение тишины, и в течение активных речевых частей.

Основываясь на наборе найденных задержек, PESQ сравнивает первоначальный (входной) сигнал с выровненным ухудшенным выходом проверяемого устройства, используя интуитивную модель. Суть этого процесса - преобразование исходного и ухудшенного сигналов к внутреннему представлению, которое является аналогичным психофизическому представлению сигналов звуковой частоты в человеческой акустической системе, принимая во внимание шумовые частоты и громкость. Это достигается в несколько этапов: временное выравнивание, выравнивание уровня к калиброванному уровню восприятия, частотно-временное соответствие, частотное изменение масштаба, сжимающее масштабирование громкости.

Точность соответствия между оценкой PESQ и субъективной оценкой может быть измерена путем вычисления коэффициента корреляции. Обычно это осуществляется на средних значениях оценки после отображения объективных оценок в субъективные. Коэффициент корреляции вычисляется с помощью формулы Пирсона:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Из-за того, что многие уровни в PESQ алгоритмически довольно сложны, описание PESQ невозможно выразить в математических формулах.

В алгоритме PESQ учитываются следующие причины ухудшения качества сигнала: искажение его при кодировании, ошибки при передаче, потеря пакетов, время задержки передачи пакетов и флуктуация этого времени, фильтрация сигнала в аналоговых сетевых компонентах.

Порядок выполнения работы

1. Запустить среду программирования Matlab R2017a.

2. Используя программу лабораторной работы № 1, изменить значения m , N и formants , а также редактировать полосовой фильтр на выходе для 3 - 5 различных аудиофайлов.

3. Сравнить полученные аудиофайлы субъективным методом, дав им оценку от 1 до 5. Результаты записать в таблицу.

4. С помощью программы PESQ сравнить первоначальный аудиофайл и аудиофайл с измененными значениями m , N , formants и полосового фильтра. Результаты записать в таблицу. Построить зависимости качества от параметров.

5. Изменяя значения m , N и formants , а также редактируя полосовой фильтр на выходе, добиться наилучшего качества речевого сигнала.

Примечание: при выполнении лабораторной работы необходимо изменить каждый параметр не менее 5 раз. Зависимости строить для усредненного значения каждого параметра.

Контрольные вопросы

1. Частотное маскирование.
2. Временное маскирование.
3. Критические полосы.
4. Субъективный метод оценки качества речевых сигналов.
5. Объективный метод оценки качества речевых сигналов.
6. Психоакустический метод оценки качества речевых сигналов.
7. Оценка PESQ.

Содержание отчета

1. Цель работы.
2. Выбранные значения параметров.
3. Результаты субъективной оценки.
4. Результаты оценки с использованием программы resq.
5. Графическое отображение результатов оценки с применением программы PESQ.

6. Значения параметров с наилучшим качеством речевого сигнала.

7. Выводы по работе.

Библиографический список

1. Берко Г. А., Галич С. В., Пасюк А. О., Семенов Е. С. Применение алгоритма PESQ для оценки качества передачи речи по IP-сетям // *Отарёв-Online*. 2015. № 11. С. 52.
2. Мешеряков Р. В. Система оценки качества передаваемой речи // Доклады ТУСУР. 2010. № 2-1. С. 22.
3. Давыдов Г. В., Каван Д. М., Шамгин Ю. В. Оценка разборчивости речи в зашумленном помещении // Доклады БГУИР. 2012. № 4. С. 66.

Лабораторная работа № 3 Кодирование речевых сигналов посредством вейнлет-преобразования

Цель работы: исследование принципов кодирования сигналов посредством вейнлет-преобразования.

Теоретическая часть

Определение вейнлет-преобразования. Основная область применения вейнлетных преобразований – анализ и обработка сигналов и функций, нестационарных во времени или неоднородных в пространстве. При этом результаты анализа должны содержать не только общую частотную характеристику сигнала (распределение энергии сигнала по частотным составляющим), но и сведения об определенных локальных координатах, на которых себя проявляют те или иные группы частотных составляющих или на которых происходят быстрые изменения. Преобразование обеспечивает частотно-временное представление сигналов.

Как правило, вейнлет-преобразования (ВП) подразделяют на дискретное (ДВП) и непрерывное (НВП). ДВП используется для преобразования и кодирования сигналов, НВП – для анализа сигналов. Вейнлеты представляют собой особые функции в виде коротких волн (всплесков) с нулевым интегральным значением и с локализацией по оси независимой переменной (t или x), способных к сдвигу по этой оси и масштабированию (растяжению/сжатию). Любой из наиболее часто используемых типов вейнлетов порождает полную ортогональную систему. В случае вейнлет-анализа (декомпозиции) процесса (сигнала) благодаря изменению масштаба вейнлеты способны выявить различие в характеристиках процесса на различных шкалах, а посредством сдвига можно проанализировать свойства процесса в различных точках на всем исследуемом интервале. Именно благодаря свойству полноты этой системы можно осуществлять восстановление (реконструкцию или синтез) процесса посредством обратного ВП.

В последние несколько лет интерес к ВП у нас резко возрос. Появились учебные пособия, монографии, переведены на русский язык фундаментальные теоретические книги И. Добеши и Т. Чуй, вейнлетам посвящены разделы и главы в учебниках.

Вейнлет-преобразование (ВП) одномерного сигнала – это его представление в виде обобщенного ряда или интеграла Фурье по системе базисных функций:

$$\psi_{a,b}(t) = \psi\left(\frac{b-t}{a}\right), \quad (3.1)$$

сконструированных из обладающего определенными свойствами математического (исходного) вейнлета $\psi(t)$ за счет операций сдвига во времени (a) и изменения временного масштаба.

Функция $\psi(t)$ называется *вейнлетом* (по-английски – *wavelet*; в русской математической литературе используется также термин *всплеск*), если:

- А. $\psi(t)$ непрерывна;
- В. $\psi(t)$ интегрируема на всей прямой;
- С. $\int_{-\infty}^{\infty} \psi(t) dt = 0$.

Вейнлет-преобразованием $f(t)$ называется функция двух переменных (**непрерывное вейнлет-преобразование**).

$$Wf(b,a) = \frac{1}{a} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{b-t}{a}\right) dt. \quad (3.2)$$

Итак, в отличие от преобразования Фурье вейнлет-преобразование определено неоднозначно: каждому вейнлету соответствует свое преобразование. Условие С означает, что Фурье-образ $\hat{\psi}(\omega)$ вейнлета обращается в 0 при $\omega = 0$. Это нужно для того, чтобы в Фурье-области вейнлет был локализован вокруг некоторой ненулевой частоты ω_0 . В качестве анализирующих вейнлетов обычно выбираются функции, хорошо локализованные также и в «пространственной области» (т.е. по t).

Дискретное вейнлет-преобразование. В принципе при

обработке данных на ПК может выполняться дискретизированная версия непрерывного вейвлет-преобразования с заданием дискретных значений параметров (a, b) вейвлетов с произвольным шагом Δa и Δb. В результате получается избыточное количество коэффициентов, намного превосходящее число отсчетов исходного сигнала, которое не требуется для реконструкции сигналов.

Дискретное вейвлет-преобразование (ДВП) обеспечивает достаточно информации как для анализа сигнала, так и для его синтеза, являясь вместе с тем экономным по числу операций и по требуемой памяти. ДВП оперирует с дискретными значениями параметров a и b, которые задаются, как правило, в виде степенных функций:

$$a = a_0 \cdot 2^m, \quad b = k \cdot a_0 \cdot 2^m, \quad a_0 > 1, \quad m, k \in \mathbb{Z},$$

где l – пространство целых чисел $\{-\infty, \infty\}$, m – параметр масштаба, k – параметр сдвига. Базис пространства $L^2(\mathbb{R})$ в дискретном представлении:

$$\psi_{mk}(t) = |a_0|^{m/2} \psi(a_0^m t - k), \quad m, k \in \mathbb{Z}, \quad \psi(t) \in L^2(\mathbb{R}). \quad (3.3)$$

Вейвлет-коэффициенты прямого преобразования:

$$C_{mk} = \int_{-\infty}^{\infty} s(t) \psi_{mk}(t) dt. \quad (3.4)$$

Значение 'a' может быть произвольным, но обычно принимается равным 2, при этом преобразование называется *двухкратным вейвлет-преобразованием*. Для двукратного преобразования разработан быстрый алгоритм вычислений, аналогичный быстрому преобразованию Фурье, что предопределило его широкое использование при анализе массивов цифровых данных.

Обратное дискретное преобразование для непрерывных сигналов при нормированном ортогональном вейвлетном базисе пространства:

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{mk} \psi_{mk}(t). \quad (3.5)$$

На рис. 3.1 изображен типичный график вейвлета и его Фурье-образа.

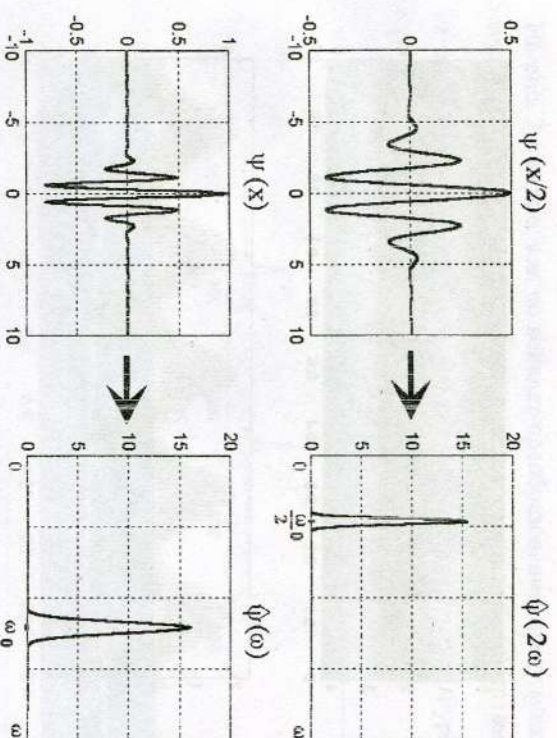


Рис. 3.1. Вейвлет в пространственной и частотной областях

Следующие рисунки (рис. 3.2 и рис. 3.3) показывают, какую информацию о сигнале можно получить при помощи вейвлет-преобразования. Вертикальная ось отвечает «масштабной» переменной a [на рисунках указаны значения, равные $\log_2(a)$, так что внизу находятся точки, отвечающие малым масштабам, а сверху – большим]. По горизонтальной оси – переменная b (совпадающая с t). Условными цветами показано распределение абсолютных значений $|W(b, a)|$. Рассмотрим сначала простейшие сигналы. «Чистым гармоникам» вида $e^{ia't}$ соответствуют яркие горизонтальные полосы, где модуль $|W(b, a)|$ велик. Локальным особенностям (нарушениям гладкости) отвечают вертикальные полосы, выходящие из точки, где находится особенность. Вейвлет-преобразование позволяет определить характер изменения частоты сигнала во времени.

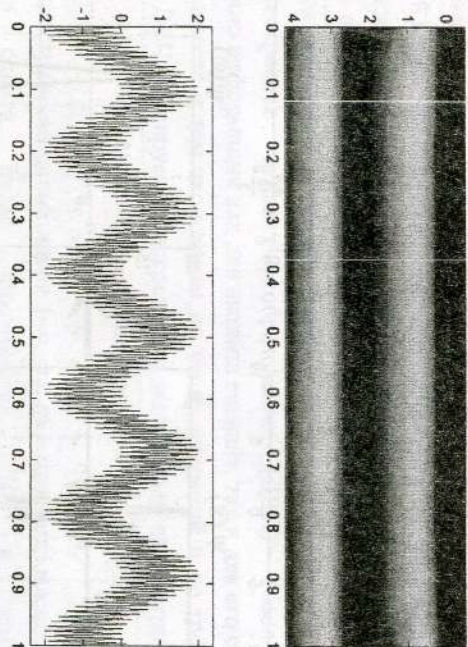


Рис. 3.2. График суммы двух гармоник (внизу);
его вейвлет-преобразование (вверху)

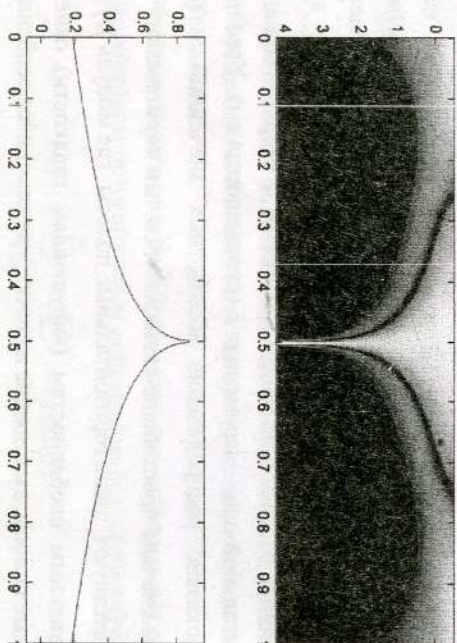


Рис. 3.3. Изолированная гильбертова сингулярность (внизу); ее
вейвлет-преобразование (вверху)

На рис. 3.4 видно, как по вейвлет-преобразованию сразу можно точно локализовать точку перехода с квадратичного «визга» (chirp) вида $e^{i\alpha t^2}$ на чистую гармонику $e^{i\beta t}$. Для сравнения приведен спектр Фурье, на котором вся эта динамика совершенно не отражена.

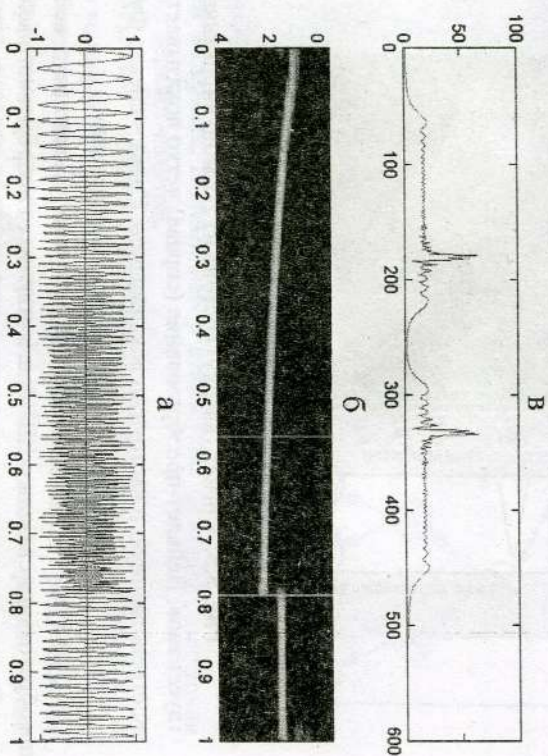


Рис. 3.4 Сигнал: «визг» (chirp) с квадратичным нарастанием
частоты мгновенно сменяется чистой гармоникой (а); вейвлет-
преобразование (б); спектр Фурье (в)

На рис. 3.5 приведен другой пример сегментации различных частотных компонент с помощью вейвлет-преобразования (здесь по вертикальной оси отложены значения нормированной частоты).

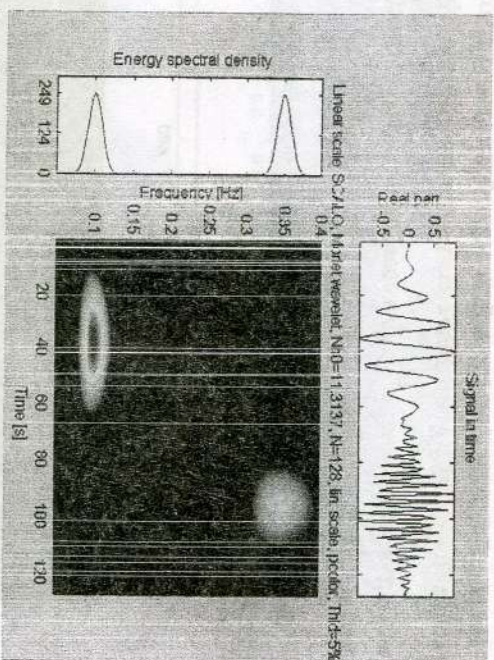


Рис. 3.5. Сигнал (вверху) состоит из 2-х модулированных гауссианов. Вейвлет-преобразование (справа) четко показывает их пространственную и частотную локализацию. Слева – спектр Фурье, дающий только частотную локализацию

Пакетные вейвлет-преобразования. Пакетные вейвлет-преобразования связаны с ДВП, т.е. в отличие от непрерывного вейвлет-преобразования имеют способность кодировать сигнал. В обычном алгоритме Маллата быстрого вейвлет-преобразования (БВП) при переходе с масштабного уровня m на уровень $m+1$ функция аппроксимирующих коэффициентов $c_{m,k}$ разделяется на низкочастотную ($c_{m+1,k}$) и высокочастотную ($d_{m+1,k}$) части спектрального диапазона, и при дальнейшем увеличении масштабных уровней аналогичному разложению последовательно подвергаются только низкочастотные функции (аппроксимирующие). В пакетном алгоритме БВП операция последовательного частотного расширения применяется как для низкочастотных, так и для высокочастотных (детализирующих) коэффициентов. В результате возникает дерево расширения, пример которого (в предельной форме расширения на всех уровнях) показан на рис. 3.6. При таком расширении вейвлеты каждого последующего уровня образуются из вейвлета предыдущего

уровня разделением на два новых вейвлета:

$$\psi_1(t) = \sum_n h_n \psi(t-n), \quad \psi_2(t) = \sum_n g_n \psi(t-n). \quad (3.6)$$

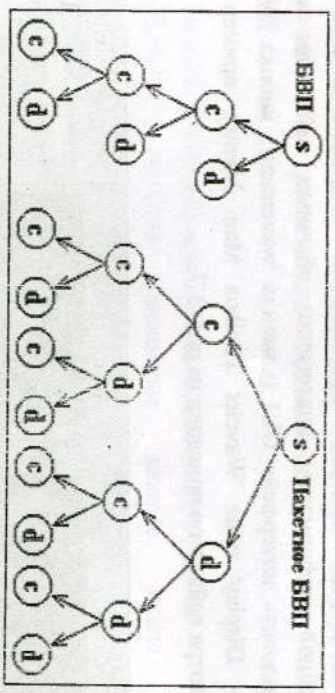


Рис. 3.6. Дерево расширения

Новые вейвлеты также локализованы в пространстве, но на вдвое более широком интервале. Полный набор вейвлетных функций разложения называют вейвлет-пакетом.

Пакетное вейвлет-преобразование является адаптивным и широко используется для компрессии сигналов и их очистки от шумов. Оно позволяет более точно приспособиваться к особенностям сигналов путем выбора соответствующей оптимальной формы дерева разложения, которая обеспечивает минимальное количество вейвлет-коэффициентов при заданной точности реконструкции сигнала и тем самым целенаправленно исключает из обратного БВП незначимые, информационно избыточные или ненужные детали сигналов. Мерой оптимальности обычно служит концентрация числа вейвлет-коэффициентов для реконструкции сигнала с заданной точностью (погрешностью). Оценка информативности набора коэффициентов выполняется по энтропии, под которой обычно понимается величина:

$$E = \exp(-\sum_n p_n \log(p_n)), \quad p_n = |x_n|^2 / \|x\|^2. \quad (3.7)$$

Любое усреднение коэффициентов увеличивает энтропию. При анализе дерева вычисляется энтропия узлов и его разделенных частей с и d. Если при разделении узла энтропия не уменьшается, то дальнейшее ветвление с этого узла не имеет смысла.

Нижеследующее описание вейджет-преобразований базируется на пакете расширения Wavelet Toolbox в системе MatLab. Для просмотра формы пакетных вейджетов и получения о них более подробной информации аналогично обычным вейджетам можно использовать интерфейс GUI (команда 'wavemenu', кнопка 'Wavelet Packet Display' окна 'Wavelet Toolbox Main Menu'), пример окна просмотра вейджетов приведен на рис. 3.7.

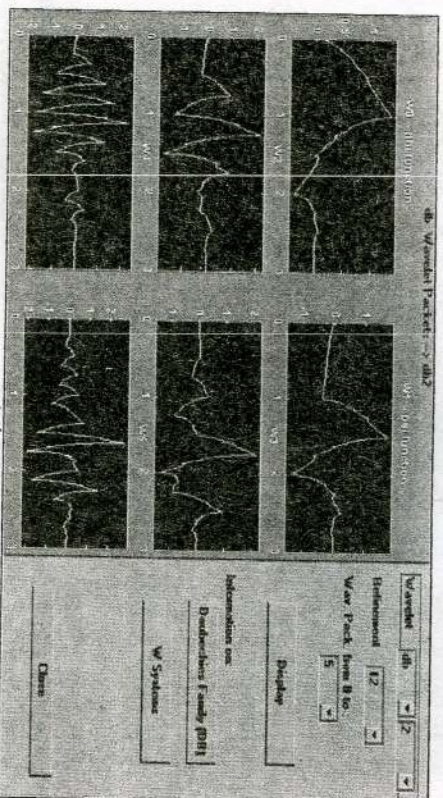


Рис. 3.7. Окно просмотра вейджетов

В качестве исследуемых базисов предложены:

1. Базисы Добеши: 'db1...db12'.
2. Биортогональные базисы: 'bior1.1', 'bior1.3', 'bior1.5', 'bior2.2', 'bior2.4', 'bior2.6', 'bior2.8', 'bior3.1', 'bior3.3', 'bior3.5', 'bior3.7', 'bior3.9', 'bior4.4', 'bior5.5', 'bior6.8'.
3. Обратные биортогональные базисы: 'rbior1.1', 'rbior1.3', 'rbior1.5', 'rbior2.2', 'rbior2.4', 'rbior2.6', 'rbior2.8', 'rbior3.1', 'rbior3.3', 'rbior3.5', 'rbior3.7', 'rbior3.9', 'rbior4.4', 'rbior5.5', 'rbior6.8'.
4. Симлеты: 'sym2', ..., 'sym8'.
5. Базисы Койфмана: 'coif1', ..., 'coif5'.

Порядок выполнения работы

Работа состоит из трех частей: исследование непрерывного вейджет-преобразования; исследование свойств двухуровневого вейджет-пакетного разложения; исследование свойств базисной функции.

1. Исследование непрерывного вейджет-преобразования

Введите заданные исходные данные, после чего раскомментируйте 9-ю строку и закомментируйте 11-ю. Получите:

```
clear 'all'; clc;
%----- Ввод исходных данных -----
signal_n = 2;
basis = 'bior2.6';
%-----
[t, s] = signal(signal_n);
plot(t,s), grid;
section1;
```

Запустите на выполнение и зарисуйте сигнал.

Раскомментируйте 11-ю строку и закомментируйте 9-ю.

Запустите на выполнение и зарисуйте вейджет-спектр (рис. 3.8).

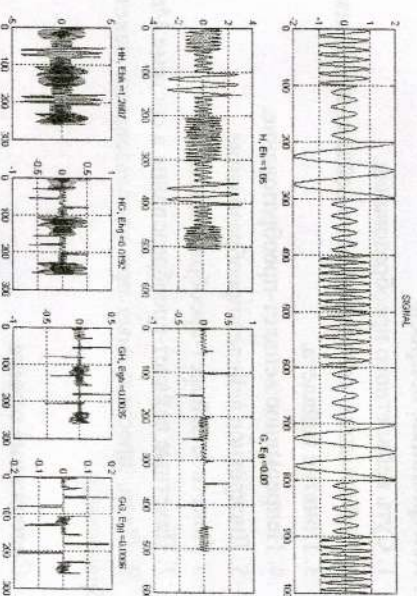


Рис. 3.8. Вейджет-спектр

- Исследование свойств двухуровневого вейлет-пакетного разложения.

Измените section1 на section2. После этого в соответствии с заданием введите первый базис для исследования. Запустите на выполнение.

Меняя базис согласно заданию, заполните таблицу:

Результаты эксперимента

Базис	Eh	Eg	Ehh	Ehg	Egh	Egg

- Постройте зависимости E_{hh} , E_{hg} , E_{gh} , E_{gg} от базиса.
- Исследование свойств базисной функции.
- Замените section2 на section3 и зарисуйте скейлинг и вейлет-функции, а также их частотные характеристики.

Контрольные вопросы

- Суть вейлетного преобразования.
- Информация, содержащаяся в вейлет-преобразовании.
- Понятие вейлета.
- Непрерывное вейлет-преобразование.
- Дискретное вейлет-преобразование.
- Пакетное вейлет-преобразование.
- Пакетные вейлет-преобразования в системе Matlab.
- Цели, преследуемые при использовании пакетного вейлет-преобразования.

Содержание отчета

- Цель работы.

- Изображение графиков, построенных в ходе выполнения работы.
- Описание полученных результатов и выводы.

Библиографический список

- Новиков И.М., Стечкин С.Б. Основные конструкции вейлскалов// Фундаментальная и прикладная математика. Т. 3. Вып. 4. 1997. С.999-1028.
- Mallat S. G. A Wavelet Tour of Signal Processing, 1998.
- Jawerth В., Sweldens W. An overview of wavelet based multiresolution analyses // SIAM Review. V. 36. 1994. P. 377-412.
- Вейлскал: подборка популярных статей о теории и приложениях вейлет-анализа // Компьютерра, № 8 (236). 2 марта 1998. С. 28-53.

Лабораторная работа № 4 Использование вейвлет-преобразования для обработки речевых сигналов

Цель работы: исследование принципов обработки речевых сигналов посредством вейвлет-преобразований.

Теоретическая часть

Вейвлет-преобразование. В основе Фурье-анализа лежит утверждение, что любую 2π -периодичную функцию можно разложить на составляющие, т.е. может быть получена суперпозиция целочисленных растяжений базисной функции e^{ix} .

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \quad (4.1)$$

где c_n — коэффициенты Фурье:

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx. \quad (4.2)$$

Процесс разложения функции проиллюстрирован на рис.4.1

Преобразование Фурье дает спектральную информацию о сигнале и описывает его поведение в частотной области:

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt. \quad (4.3)$$

При переходе в частотную область полностью теряется информация о времени, что делает непригодным метод спектрального анализа при обработке нестационарных сигналов, в которых определяющее значение имеет момент времени, в который произошло то или иное событие.

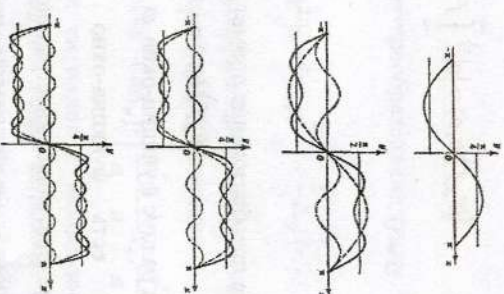


Рис. 4.1. Разложение функции

В отличие от кратковременного преобразования Фурье, которое обеспечивает равномерную сетку в частотно-временной области, вейвлет-преобразование имеет неравномерное разрешение, что позволяет исследовать сигнал как локально, так и полностью.

Так как частота обратно пропорциональна периоду, то требуется более узкое окно для локализации высокочастотной составляющей сигнала и более широкое для низкочастотной составляющей. Кратковременное преобразование Фурье допустимо применять для сигнала со сравнительно узкой полосой частот. Для широкополосного сигнала хотелось бы иметь окно, способное изменять свою ширину при изменении частоты.

Введем функцию $\varphi \in L^2(\mathbb{R})$, удовлетворяющую условию

$$\int_{-\infty}^{\infty} \frac{|\hat{\varphi}(\omega)|^2}{|\omega|} d\omega < \infty, \quad (4.4)$$

и назовем ее «базисным вейвлетом».

Относительно каждого базисного вейвлета интегральное вейвлет-преобразование определяется как

$$(W_{\varphi} f)(b, a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \varphi\left(\frac{t-b}{a}\right) dt, \quad (4.5)$$

где $a, b \in \mathbb{R}$, $a \neq 0$.

Обозначим

$$\varphi_{b,a}(t) = |a|^{-\frac{1}{2}} \varphi\left(\frac{t-b}{a}\right). \quad (4.6)$$

Интегральное преобразование примет вид

$$(W_{\varphi} f)(b, a) = \langle f, \varphi_{b,a} \rangle. \quad (4.7)$$

Если центр и радиус функции-окна φ соответственно равны t^* и Δ_{φ} , то $\varphi_{b,a}(t)$ есть функция-окно с центром $b+at^*$ и радиусом $a\Delta_{\varphi}$. Следовательно, интегральное вейвлет-преобразование локализует аналоговый сигнал во временном окне:

$$[b+at^* - a\Delta_{\varphi}, b+at^* + a\Delta_{\varphi}]. \quad (4.8)$$

Рассмотрим

$$\frac{1}{2\pi} \hat{\varphi}_{b,a}(\omega) = \frac{|a|^{-\frac{1}{2}}}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \varphi\left(\frac{t-b}{a}\right) dt = \frac{a|a|^{-\frac{1}{2}}}{2\pi} e^{-i\omega a} \hat{\varphi}(a\omega). \quad (4.9)$$

Пусть центр и радиус функции-окна $\hat{\varphi}$ равны соответственно ω^* и $\Delta_{\hat{\varphi}}$.

Тогда сместим центр окна на ω^* в 0 и обозначим

$$\eta(\omega) = \hat{\varphi}(\omega + \omega^*). \quad (4.10)$$

Применим равенство Парсеваля:

$$(W_{\varphi} f)(b, a) = \frac{a|a|^{-\frac{1}{2}}}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega b} \overline{\eta\left(a\left(\omega - \frac{\omega^*}{a}\right)\right)} d\omega. \quad (4.11)$$

Очевидно, что окно

$$\eta\left(a\left(\omega - \frac{\omega^*}{a}\right)\right) = \eta(a\omega - \omega^*) = \hat{\varphi}(a\omega) \quad (4.12)$$

имеет радиус $\frac{1}{a} \Delta_{\hat{\varphi}}$.

Интегральное вейвлет-преобразование также локализует сигнал по частоте с окном

$$\left[\frac{\omega^*}{a} - \frac{1}{a} \Delta_{\hat{\varphi}}, \frac{\omega^*}{a} + \frac{1}{a} \Delta_{\hat{\varphi}}\right]. \quad (4.13)$$

Аналогично преобразованию Габора введем частотно-временное окно для интегрального вейвлет-преобразования:

$$[b+at^* - a\Delta_{\varphi}, b+at^* + a\Delta_{\varphi}] \times \left[\frac{\omega^*}{a} - \frac{1}{a} \Delta_{\hat{\varphi}}, \frac{\omega^*}{a} + \frac{1}{a} \Delta_{\hat{\varphi}}\right]. \quad (4.14)$$

Обработка речевых сигналов с использованием вейвлет-преобразования. В последние годы вейвлеты находят широкое применение при фильтрации, предварительной обработке и синтезе различных сигналов, решении задач сжатия и обработки изображений, анализе состояния и прогнозирования ситуации на фондовых рынках, а также во многих других случаях.

Вейвлеты были предложены математиками и, по существу, являются новыми математическими понятиями и объектами. Особенно важна принципиальная возможность вейвлетов представлять нестационарные сигналы. Все большее число специалистов по цифровой обработке сигналов убеждается в том, что преобразование Фурье в классическом виде не обеспечивает необходимую точность представления нестационарных сигналов, в частности, к которым относятся речевые сигналы. Значительный интерес представляют утверждения о том, что вейвлет-спектрограммы намного более информативны, чем обычные Фурье-спектрограммы, и (в отличие от последних) позволяют легко выявлять тончайшие локальные особенности акустических сигналов.

Вейвлеты (wavelets) — это обобщенное название особых функций, имеющих вид коротких волновых пакетов с плавным интегральным значением и с той или иной, подчас очень сложной, формой, локализованных по оси независимой переменной (t или x) и способных

к сдвигу по ней и масштабированию (сжатию/растяжению). Вейвлеты создаются с помощью специальных базовых функций – прототипов, заданных их вид и свойства и удовлетворяющих целому ряду условий. Набор вейвлетов в их временном или частотном представлении может приближать сложный сигнал или изображение, причем идеально точно или с некоторой погрешностью. Вейвлет-преобразование (ВП) в большой степени позволяет преодолеть основные недостатки преобразования Фурье ФП, поскольку базисные функции ВП обладают свойством временной локализации, т.е. обладают конечной энергией (нормой):

$$E_f = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty \quad (4.15)$$

Порядок выполнения работы

1. Сгенерировать стационарные, нестационарные сигналы и сигналы с шумом.

Пример

$t=0:0.1:6*\pi$;

Стационарный сигнал

$y=\sin(t)$;

$z=\sin(t)+\sin(2*t)$;

Сигнал с шумом

$N=\text{rand}(1,189)$;

$w=\sin(t)$;

$w=w+N$;

Нестационарный сигнал

$t=0:0.1:2*\pi$;

$w(1:63)=\sin(t)$;

$w(64:126)=\cos(t)$;

$w(127:190)=\cos(2*t)$;

Далее их нужно сохранить (каждый сигнал в отдельном файле), для этого в окне рабочей области выделяется нужная переменная и в контекстном меню выбирается пункт *Save Selection As...* (рис. 4.2).

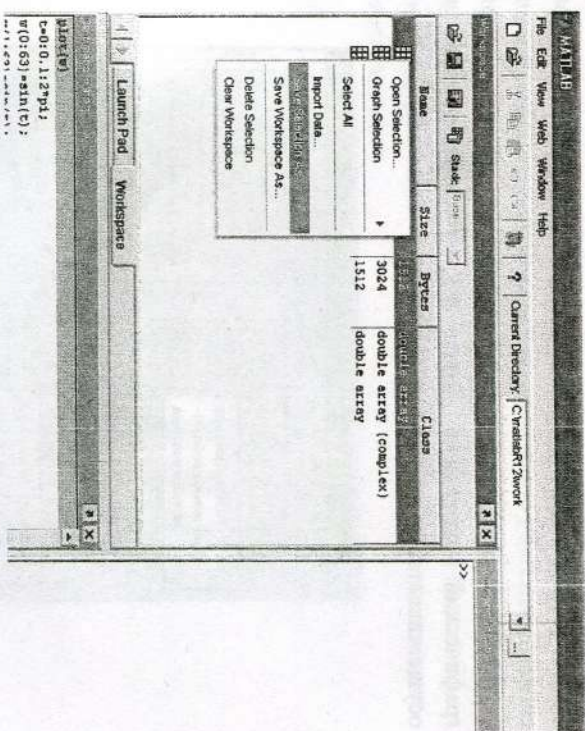


Рис. 4.2. Контекстное меню

2. Проанализировать сигналы с использованием преобразования Фурье, объяснить результаты.

Для построения Фурье-спектра используется функция $\text{fft}(\text{имя сигнала}, \text{число точек ДПФ})$

Пример

$Y=\text{fft}(\text{Sig}, 512)$

$A=\text{abs}(Y)$;

$\text{plot}(A(1:\text{length}(A)/2));$

3. Рассмотреть кратковременное преобразование Фурье для анализируемого сигнала, объяснить результаты.

Для построения спектрограммы используется функция

$\text{Spectgram}(\text{имя сигнала})$

Пример

$\text{Spectgram}(\text{Sig})$

4. Проанализировать полученные сигналы с использованием различных вейвлетов (не менее 3), объяснить результаты, определить

«оптимальный» (дающий наибольшую информацию) вейвлет для сигнала.

Для построения вейвлет-спектра можно использовать графический интерфейс (рис. 4.3, рис.4.4), вызов которого осуществляется командой waveletmenu.

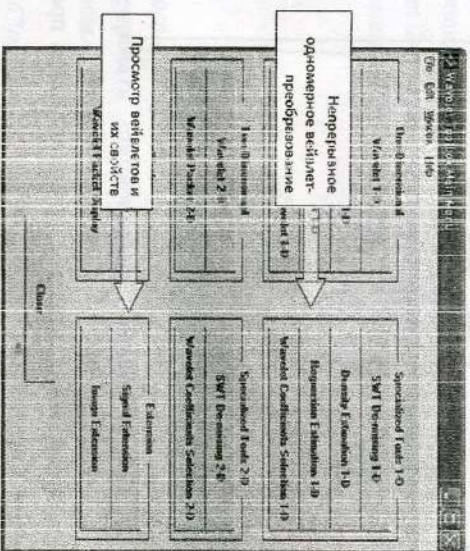


Рис. 4.3. Окно программы Wavelet Tool Box Main Menu

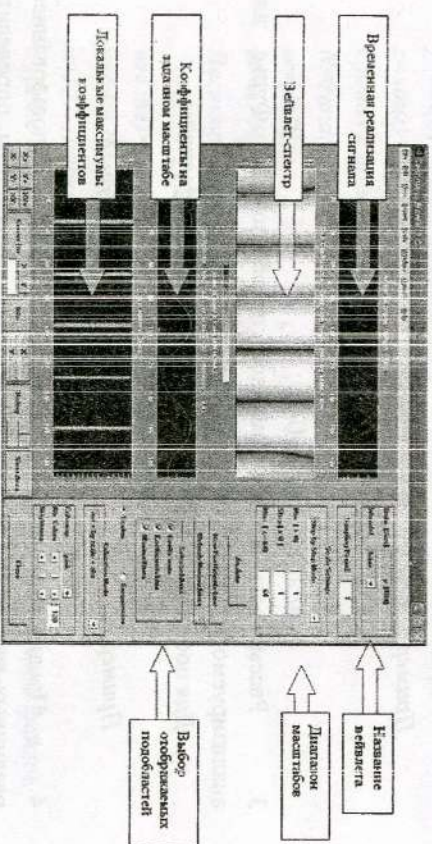


Рис. 4.4. Окно программы Continuous Wavelet

Для загрузки сигнала используется пункт меню File/Load Signal. На рис. 4.5 приведен пример анализа сигнала $x = \sin(t) + \sin(2 \cdot t)$. Видно, что сигнал содержит две частоты, разделенных на масштабе ~70. Вейвлет-коэффициенты меняются периодически, что доказывает периодичность сигнала.

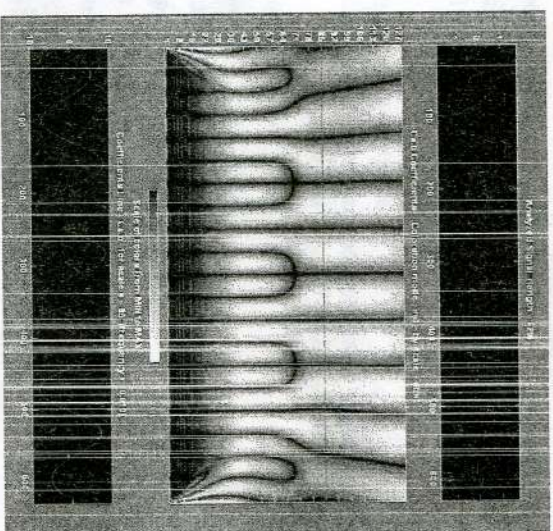


Рис. 4.5 Пример анализа сигнала

При анализе нестационарного сигнала вейвлет-спектр показывает изменение частоты в момент времени 500, а также изменение, произошедшее в момент времени ~250, причем можно сделать вывод, что частота сигнала в данном случае осталась неизменной.

5. Сравнить и объяснить результаты Фурье- и вейвлет-анализов.
 - Варианты заданий:
 - Для всех вариантов:
 - t_1, t_2, t_3, t_4 выбираются таким образом, чтобы получаемые сигналы содержали не менее 2-х периодов;

• для генерации сигнала с шумом используются стационарный сигнал и шум с равномерной плотностью распределения.

Вариант № 1

$$Y = 3 \sin(t)$$

$$w(t_1, t_2) = \sin(t); w(t_2, t_3) = \sin(5t); w(t_3, t_4) = \sin(10t);$$

Вариант № 2

$$Y = \sin(5t)$$

$$w(t_1, t_2) = \sin(t); w(t_2, t_3) = \sin(10t); w(t_3, t_4) = \sin(5t);$$

Вариант № 3

$$Y = \sin(10t)$$

$$w(t_1, t_2) = \sin(t); w(t_2, t_3) = \lg(5t); w(t_3, t_4) = \sin(10t);$$

Вариант № 4

$$Y = \sin(t) + \sin(5t)$$

$$w(t_1, t_2) = 50 \sin(t); w(t_2, t_3) = \lg(5t); w(t_3, t_4) = 50 \sin(5t);$$

Вариант № 5

$$Y = \sin(t) + \sin(10t)$$

$$w(t_1, t_2) = \text{square}(t); w(t_2, t_3) = \text{square}(10t); w(t_3, t_4) = \text{square}(5t);$$

Вариант № 6

$$Y = \sin(10t) + \sin(2t)$$

$$w(t_1, t_2) = \text{sawtooth}(t); w(t_2, t_3) = \text{sawtooth}(10t); w(t_3, t_4) = \text{sawtooth}(5t);$$

Вариант № 7

$$Y = \sin(t) + \sin(10t) + \sin(15t)$$

$$w(t_1, t_2) = \text{sawtooth}(t, 0, 5); w(t_2, t_3) = \text{sawtooth}(10t, 0, 5);$$

$$w(t_3, t_4) = \text{sawtooth}(5t, 0, 5);$$

Вариант № 8

$$Y = \sin(t) + \sin(15t)$$

$$w(t_1, t_2) = \text{sinc}(t); w(t_2, t_3) = \text{sinc}(10t); w(t_3, t_4) = \text{sinc}(5t);$$

Вариант № 9

$$Y = \sin(t) + \cos(10t)$$

$$w(t_1, t_2) = \sin(t) + \cos(5t);$$

$$w(t_2, t_3) = \sin(t) + \cos(10t); w(t_3, t_4) = \sin(t) + \cos(15t);$$

Вариант № 10

$$Y = \sin(t) + \cos(20t)$$

$$w(t_1, t_2) = \sin(15t) + \cos(5t);$$

$$w(t_2, t_3) = \sin(10t) + \cos(10t); w(t_3, t_4) = \sin(5t) + \cos(15t);$$

Контрольные вопросы

1. Отличие вейвлет-преобразования от преобразования Фурье.
2. Интегральное вейвлет-преобразование.
3. Преобразование Габора.
4. Применение вейвлетов.
5. Недостатки преобразования Фурье.

Содержание отчета

Отчет должен содержать:

- 1) временную реализацию исследуемых сигналов;
- 2) Фурье-спектры сигналов;
- 3) спектрограммы сигналов;
- 4) формы используемых вейвлетов;
- 5) вейвлет-спектры сигналов;
- 6) результаты анализа и сравнения.

Библиографический список

1. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук. 1996. Т. 166. № 11. С. 1145-1170.
2. Дремин И.М., Иванов О.В., Нечитайло В.А. Вейвлеты и их использование // Успехи физических наук. 2001. № 5. Т. 171. С. 264-301.
3. Чуи Ч. Введение в вейвлеты: пер. с англ. М.: Мир, 2001. С. 145-241.

Содержание

Лабораторная работа № 1. Изучение принципов работы вокодеров при кодировании речевых сигналов.....	1
Лабораторная работа № 2. Оценка качества речевых сигналов в МТС.....	14
Лабораторная работа № 3. Кодирование речевых сигналов по средством вейвлет-преобразования.....	26
Лабораторная работа № 4. Использование вейвлет-преобразования для обработки речевых сигналов.....	38