

Учреждение образования  
«Полоцкий государственный университет  
имени Евфросинии Полоцкой»



# ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ В ЗАДАЧАХ ОБРАБОТКИ И РАСПОЗНАВАНИЯ РЕЧИ

Бураченко Ирина Брониславовна,  
*к.т.н., доцент,  
заведующий кафедрой математики и  
компьютерной безопасности*

г. Новополоцк

## Актуальность

Обнаружение слабых речевых сигналов в шумах представляет задачу, которая имеет отношение не только при проектировании и контроле аппаратуры связи, но и систем защиты речевой информации от утечки по техническим каналам.

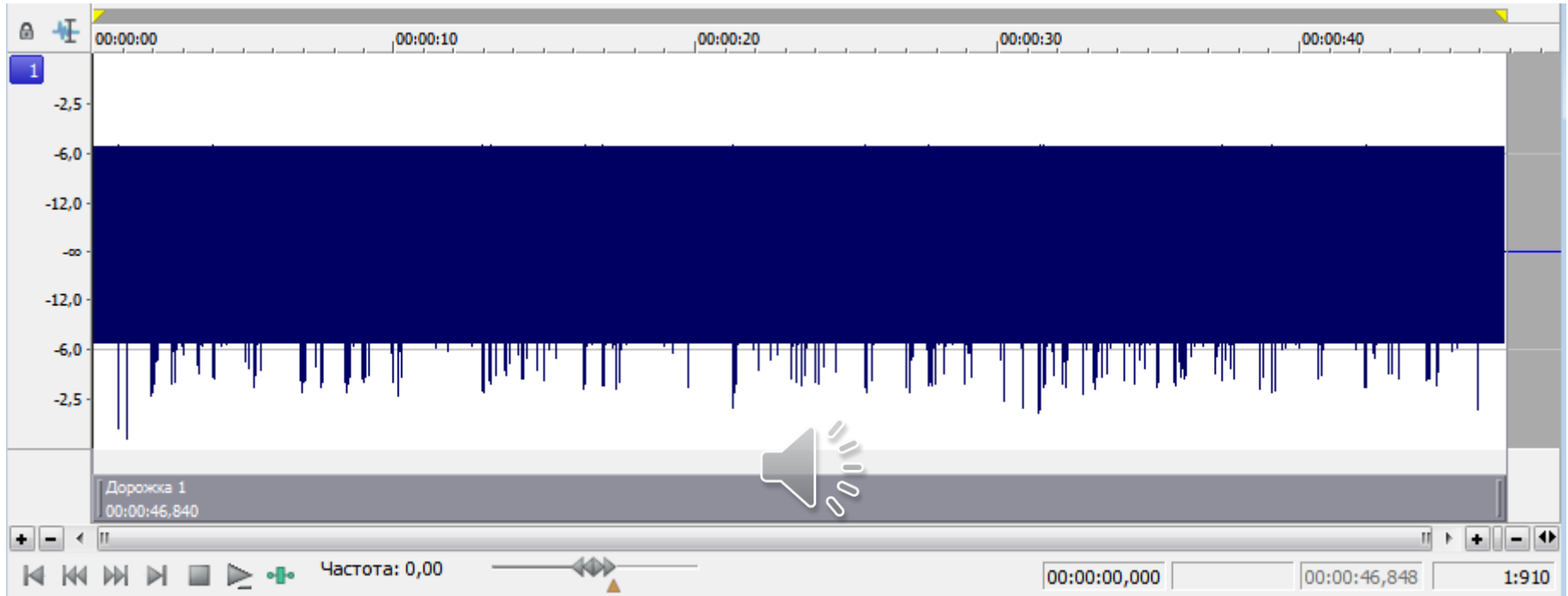
## Цель исследования

Определение метода, повышающего точность оценки информационных признаков речевого сигнала в условиях шумов.

## Решаемые задачи

1. Определение факта присутствия речевого сигнала в анализируемой смеси сигнала и маскирующего его шума выделением частоты основного тона –  $F_0$ , Гц.
2. Определение формант  $F_1, \dots, F_n$  Гц – источника получения дополнительной информации, не только о речевом сигнале, но и об индивидуальных признаках говорящего.

# ПРИМЕР МАСКИРОВАННОГО СИГНАЛА

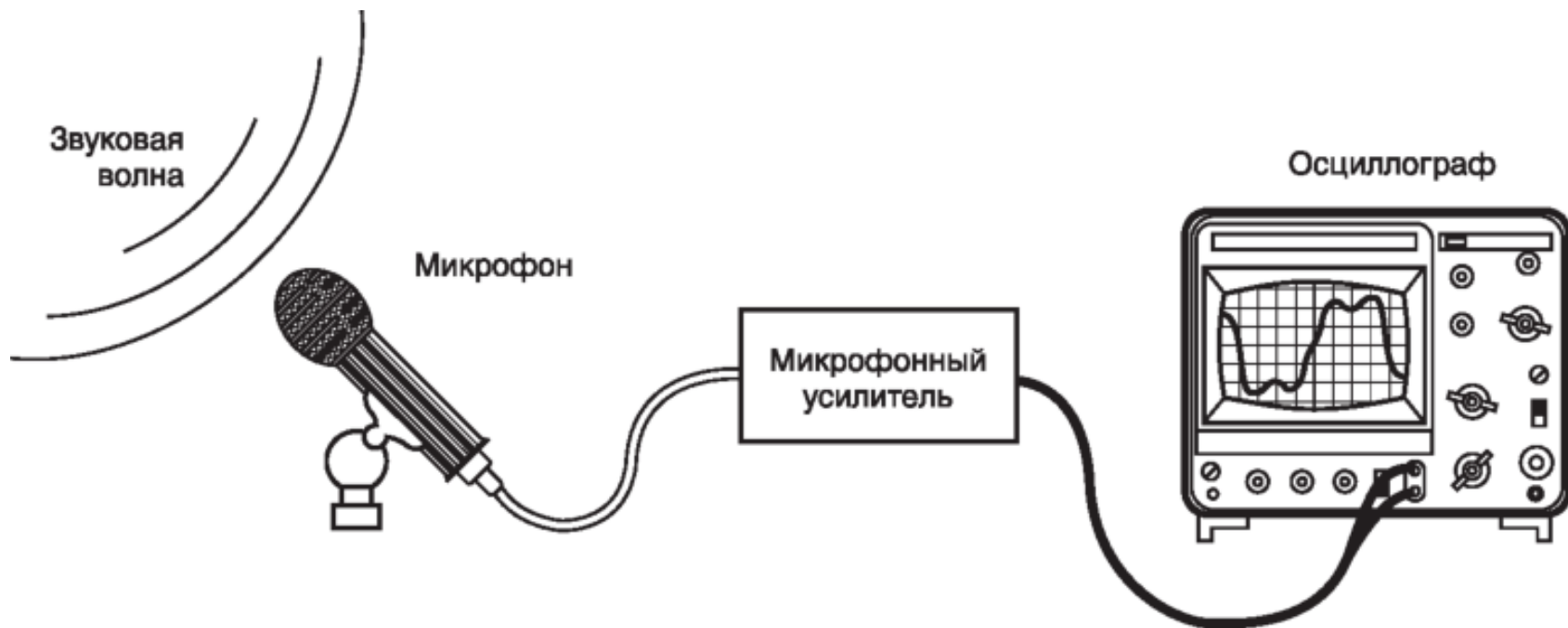


# ВВЕДЕНИЕ В АКУСТИКУ

**Звук** — это колебательное движение в любой материальной среде.

Звук характеризуется:

- амплитудой (силой)
- частотой (количеством колебаний в секунду)



# ПОНЯТИЕ СИГНАЛА

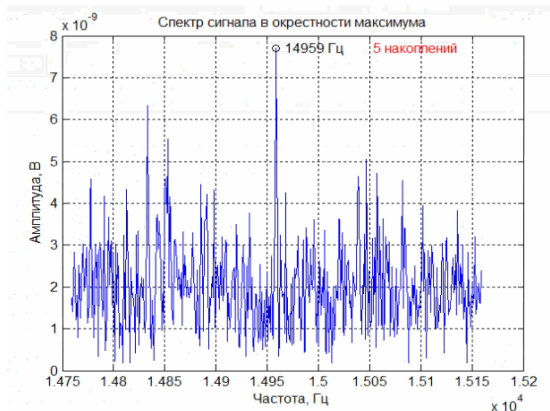
**Сигналом** называется физический процесс, параметры которого содержат информацию (сообщение) и который пригоден для обработки и передачи на расстояние.

Сигнал

электрический

акустический

оптический

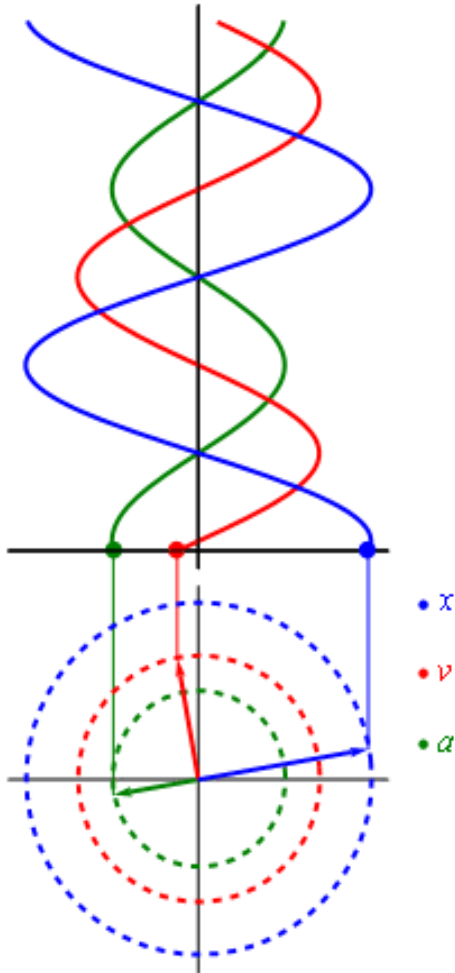


В зависимости от формы акустических колебаний различают сигналы:

- простые (тональные)
- сложные

# ПРОСТОЙ СИГНАЛ

**Простой** (тональный) — это сигнал, вызываемый колебанием, совершающимся по синусоидальному закону.



## Гармонический сигнал:

$$f(x) = A \sin(\omega t + \varphi),$$

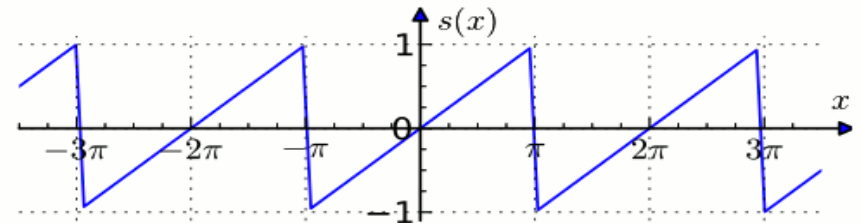
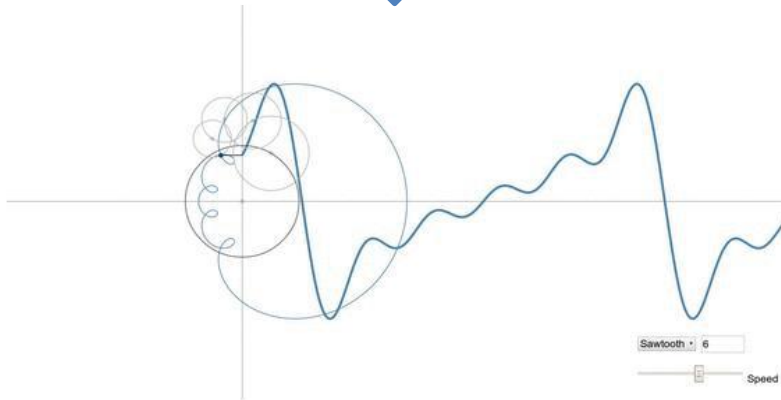
где  $A$  — длина вектора (амплитуда колебаний),  
 $\varphi$  — начальный угол (фаза) вектора в нулевой момент времени,  
 $\omega$  — угловая скорость вращения, которая равна:  
 $\omega = 2\pi f$ , где  $f$  — частота в Гц.

Зная частоту сигнала, амплитуду и угол, мы можем построить гармонический сигнал.

# СЛОЖНЫЙ СИГНАЛ

**Сложный** — это сигнал, который включает в себя целый спектр гармонических составляющих.

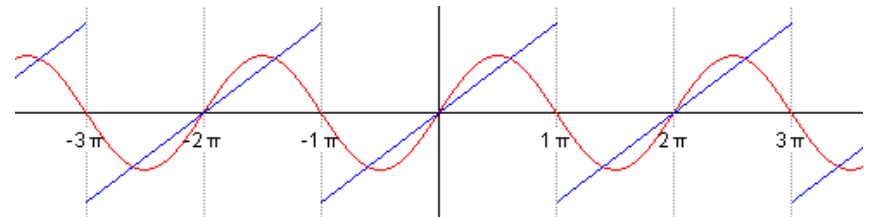
Пример пилообразного сигнала.



Его сумма будет представлена следующей формулой:

$$s(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

$$= \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(nx), \quad \text{for } x - \pi \notin 2\pi\mathbf{Z}.$$



Если мы будем поочерёдно суммировать, брать сначала  $n=1$ , затем  $n=2$  и т.д., то увидим, как у нас гармонический синусоидальный сигнал постепенно превращается в пилу.

# РЕЧЕВОЙ СИГНАЛ

Речевой сигнал имеет двойственную природу:

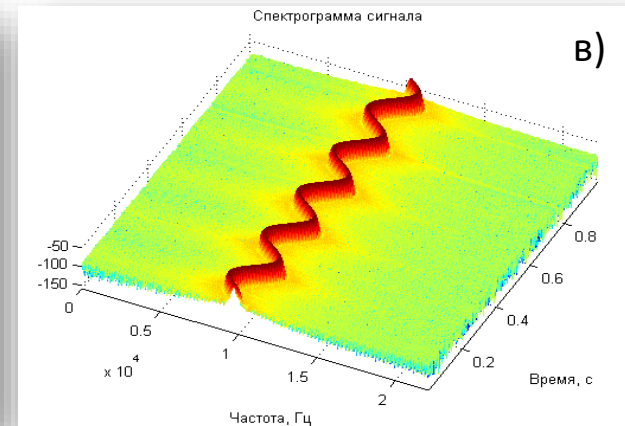
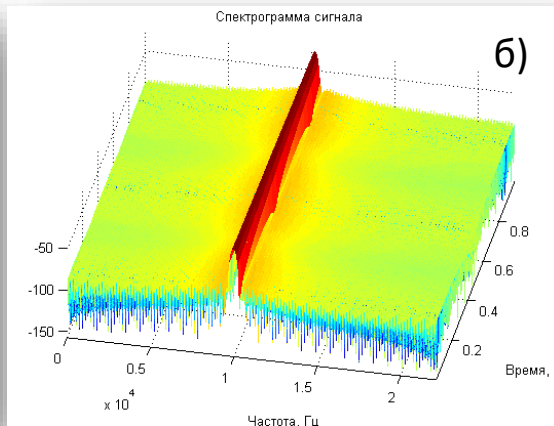
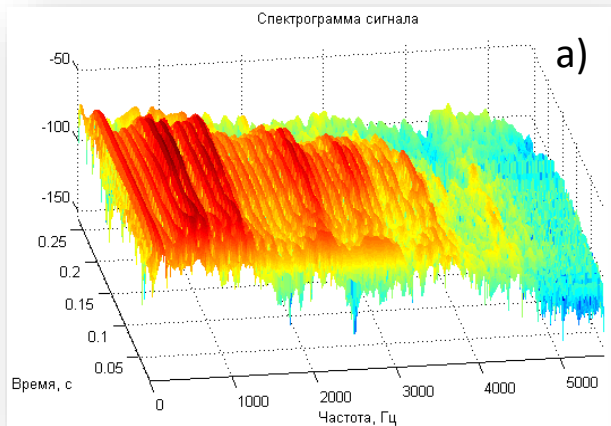
1

обычный акустический сигнал, который представляет собой процесс распространения энергии акустических колебаний в упругой среде.

2

речь как физическое явление вызывает определенные субъективные слуховые ощущения:

- *громкости,*
- *высоты,*
- *тембра,*
- *локализации,*
- *маскировки и др.*



Речевой сигнал

ЛЧМ сигнал

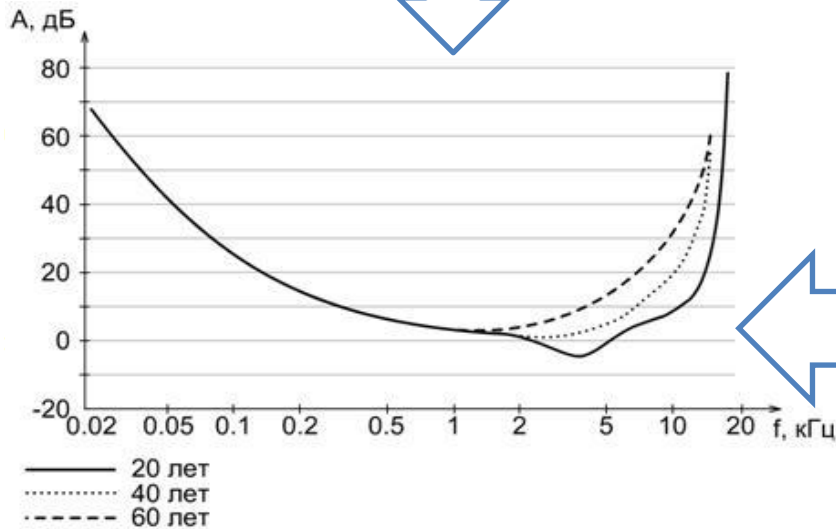
# ПОРОГ СЛЫШИМОСТИ ЗВУКОВ

Кривая порога слышимости, характеризующая наименьшую интенсивность звука определенной частоты  $f$ , который может быть услышан человеком в тишине, хорошо аппроксимируется с помощью нелинейной функции, вычисляемой по формуле:

$$G_0(f) = 3.64 \left(\frac{f}{1000}\right)^{-0.8} - 6.5 \left[ e^{-0.6 \left(\frac{f}{1000} - 3.3\right)^2} \right] + 10^{-3} \left(\frac{f}{1000}\right)^4$$

где  $G_0(f)$  – порог слышимости, дБ;  $f$  – частота звукового сигнала, Гц.

Кривая минимального порога слышимости слухового аппарата человека в тишине



Данная кривая представляет собой геометрическое место точек, которые соответствуют тональным компонентам разных частот, имеющих одинаковую громкость.

Рис. 1 Зависимость порога слышимости от частоты

# СПЕКТР РЕЧЕВОГО СИГНАЛА

Качество речи получается вполне удовлетворительным при ограничении спектра частотами **300..3400 Гц**.  
 Причем в указанной полосе частот слоговая разборчивость составляет около **90%**, разборчивость фраз - более **99%** и сохраняется **удовлетворительная натуральность звучания**.

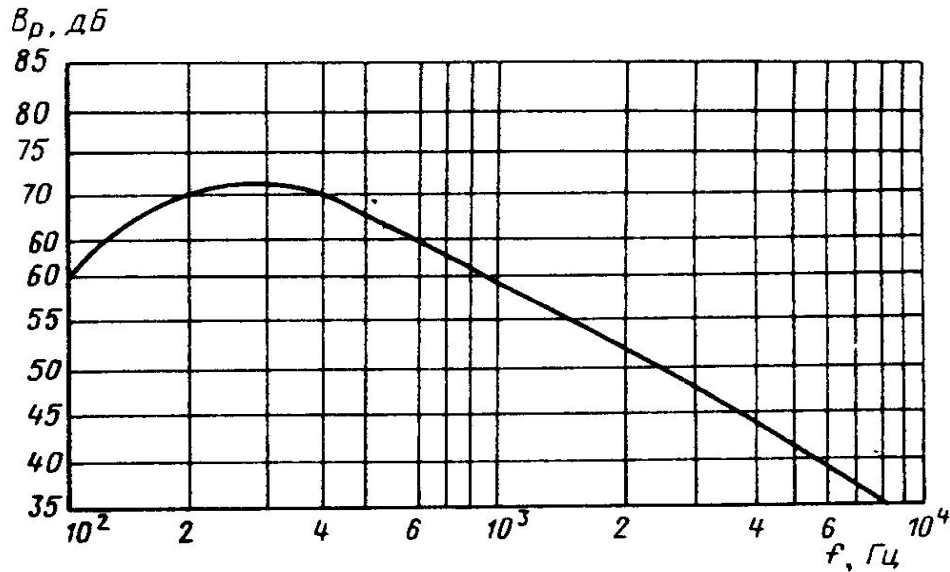
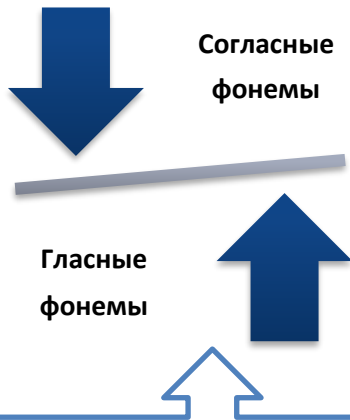


Рис. 1 Спектральная плотность речи

**Таблица** Разбиение спектра речевого сигнала на полосы равной разборчивости

Номер	Границы ( $f_1-f_2$ ), Гц	Центральная частота, $f_0$ , Гц	Ширина, $2\Delta f$ , Гц
1	100-420	260	320
2	420-570	495	150
3	570-710	640	140
4	710-865	788	155
5	865-1030	946	165
6	1030-1220	1125	190
7	1220-1410	1315	190
8	1410-1600	1505	190
9	1600-1780	1690	180
10	1780-1960	1870	180
11	1960-2140	2050	180
12	2140-2320	2230	180
13	2320-2550	2435	230
14	2550-2900	2725	350
15	2900-3300	3100	400
16	3300-3660	3480	360
17	3660-4050	3855	390
18	4050-5010	4530	960
19	5010-7250	6130	2240
20	7250-10000	8625	2750

# ОБОСНОВАНИЕ ВЫБОРА ГЛАСНЫХ ФОНЕМ



В русском языке насчитывается 42 основные и 3 неопределенные фонемы.

Звуки речи разделяют на гласные и согласные фонемы.

- Гласные звуки составляют примерно **43,5%**, а согласные – **56,5%** общего числа звуков, при этом невокализованные звуки составляют 32%. Все гласные звуки являются вокализованными.
- Гласные звуки в среднем имеют мощность **650-700 мкВт**, самые слабые согласные **0,65...0,7 мкВт**.
- Самыми редкими являются звуки: среди гласных – *э*, среди согласных – *щ*, *фь*.
- Динамический диапазон уровней речи находится в пределах 35-45 дБ.
- Гласные звуки речи имеют в среднем длительность около **0,15 с**, согласные – около **0,08** (звук *п* – около 30 мс).

Звук русской речи	Вероятность его появления в речи
О	0,078
А	0,0713
Е	0,0626
И	0,0460
У	0,0226
Я	0,0152
Ы	0,0091
Пробел /(пауза)	0,17

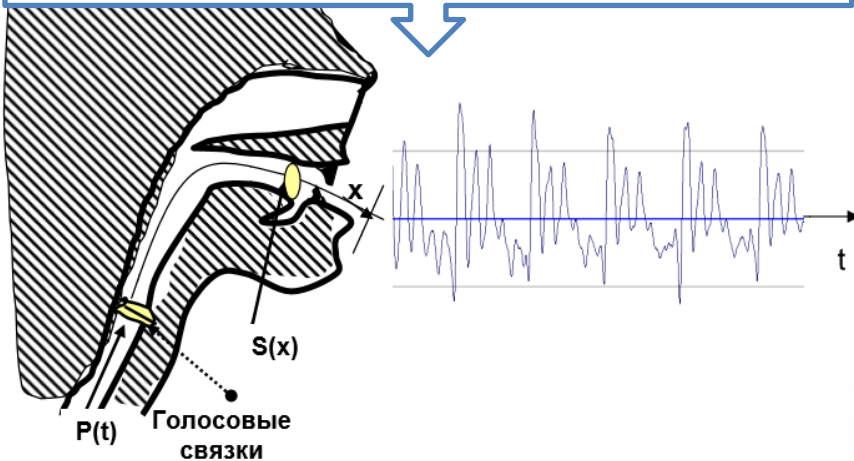
6 гласных

Звук русской речи	Вероятность его появления в речи
Н	0,0476
Р	0,0465
В	0,0418
С	0,0398
Т	0,0387
Л	0,0361
К	0,0349
М	0,0240
Д	0,0238
П	0,0210

Звук русской речи	Вероятность его появления в речи
Г	0,0168
Б	0,0120
Ь	0,0119
Й	0,0111
З	0,0104
Ж	0,0061
Ц	0,0051
Х	0,0049
Ф	0,0035
Щ	0,0012

# ФОНЕМА «А»

Схематичное изображение генерации речи



Фрагмент длительностью 0,1 с, где сосредоточены основные форманты  $f_1, f_2$ .



Исходные параметры сигнала:

**фонема:** русский язык, ударный изолированный звук «а»,

**голос:** мужской

**частота дискретизации**  $f_s = 44100$  Гц

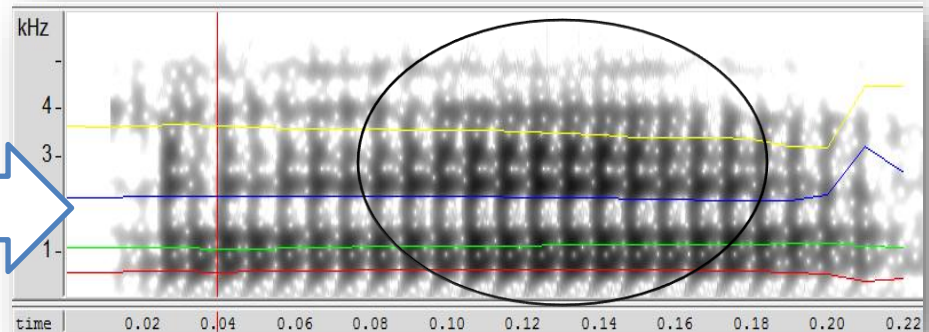
**длительность**  $t = 0,25$  с

Вычисленные параметры сигнала:

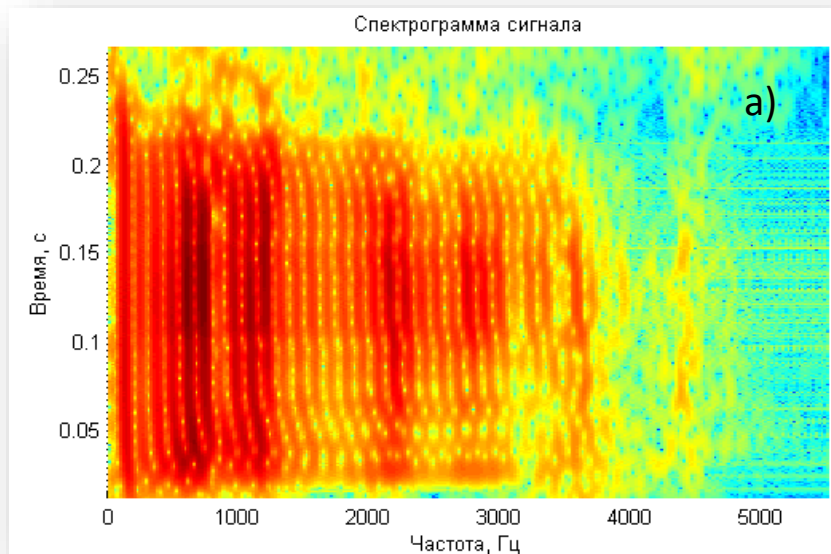
**частота основного тона:**  $f_0 = 119$  Гц

**частота первой форманты:**  $f_1 = 601$  Гц

**частота второй форманты:**  $f_2 = 1198$  Гц



## СПЕКТРОГРАММА ФОНЕМЫ «А»



Процесс речеобразования сопровождается резонансными явлениями

Собственные частоты резонаторов называют **формантами звука**, так как они формируют характерное звучание гласных и согласных.

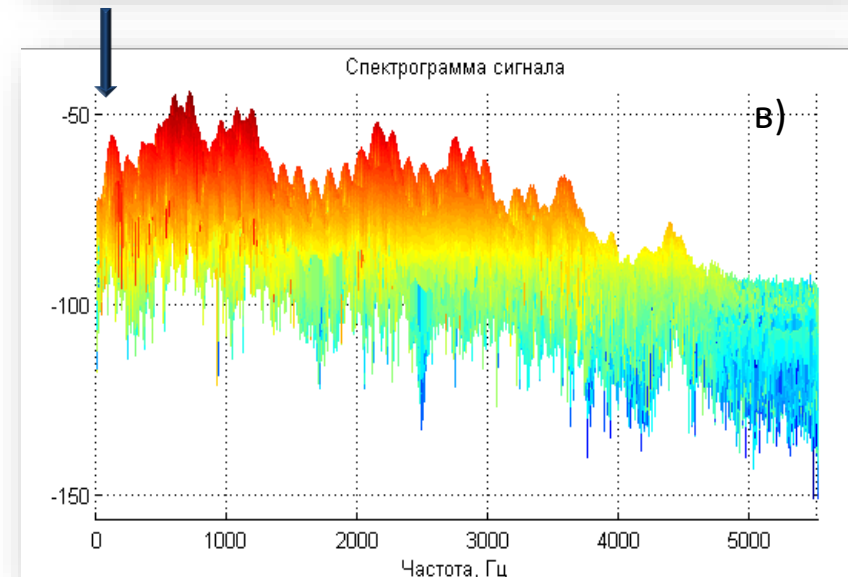
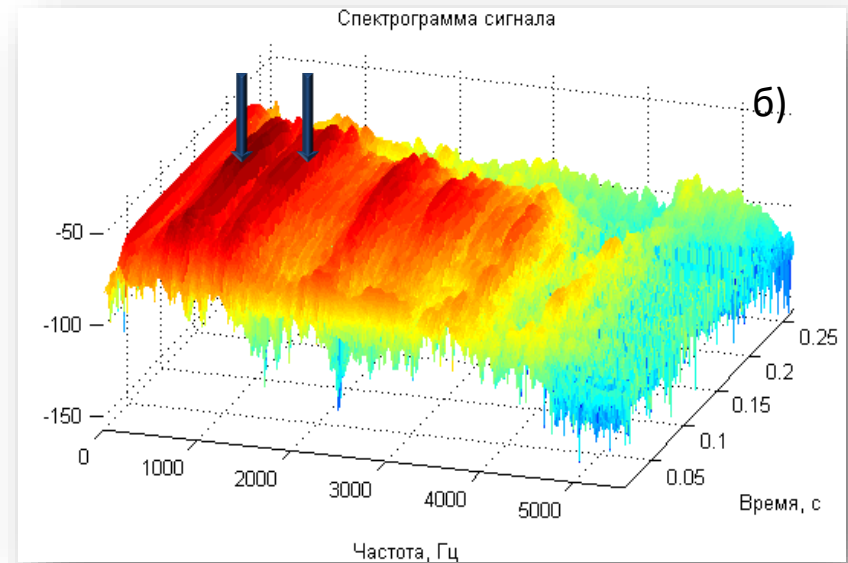


Рис. 1. Спектрограммы фонемы «А» (мужской голос)

# ОСНОВНОЙ ТОН

Одним из важнейших параметров речи, является частота основного тона для вокализованных участков.

Основную частоту колебания, соответствующую числу импульсов в единицу времени,  $F_0 = 1/T$  называют основным тоном голоса, а  $T$  – периодом основного тона.

В среднем частота основного тона составляет:



# МЕТОДЫ ОЦЕНКИ ОСНОВНОГО ТОНА

Метод

Амплитудной селекции

Спектральный

Кепстральный

Принцип

Основан на использовании расстановок меток в точках максимальных значений квазипериодических участков речевого сигнала  $s(t)$

Основан на спектральной модели стационарного сигнала:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s(t) e^{-i\omega t} dt,$$

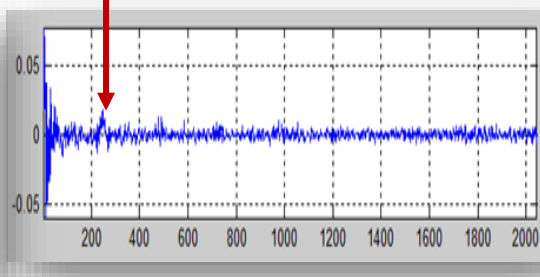
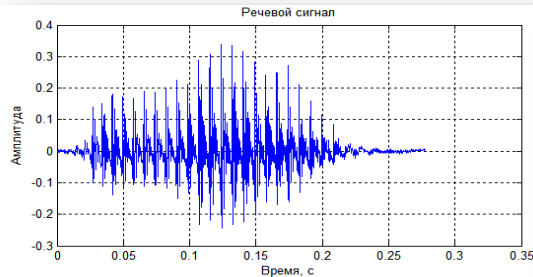
где  $s(t)$  – массив временных значений речевого сигнала;  $t$  – время;

$\omega = 2\pi/T_c = 2\pi f_0$  – круговая частота.

Основан на вычислении и анализе кепстра – обратного преобразования Фурье логарифма спектра мощности сигнала

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln|S(\omega)|^2 e^{i\omega q} d\omega,$$

где  $S(\omega)$  – спектр сигнала.



Достоинства

- прост в реализации;
- не требует больших вычислительных ресурсов.

- прост в реализации;
- не требует больших вычислительных ресурсов.

- наилучший результат этот метод дает при оценке вокализованных звуков.

Недостатки

- результаты оценки данным методом имеют низкую точность и устойчивость даже при небольших уровнях шума;
- при амплитудном ограничении высока вероятность пропуска максимума и неверного определения частоты ОТ;
- при порогах отсечки выше 30% данный расчет не представляется возможным.

- отсутствие характеристик для шумовых составляющих в произносимых согласных звуках;
- поиск максимума необходимо производить в интервале 50-350 Гц, однако в указанной полосе частот часто находится и вторая гармоника, иногда даже с большей энергией, и в этом случае она может быть ошибочно принята за частоту ОТ;
- необходимо использование спектральных окон (весовой функции).

- работа ведется в нереальном масштабе времени;
- для повышения точности оценки необходимо применять временные окна и операции сглаживания;
- имеет невысокую стойкость к шумам;
- сложен в вычислениях.

Метод

Линейного предсказания  
LPC (Linear Predictive Coding)

Автокорреляционный

На основе вейвлет-преобразования

Принцип

Основан на оценке спектра сигнала на выходе линейного тракта с неизвестными параметрами

$$H(z) = \frac{\sum_{i=0}^{N-1} b_i z^{-i}}{1 + \sum_{k=1}^{M-1} a_k z^{-k}}$$

Основан на оценке периодичности сигнала в зависимости от его задержки  
Для анализа речевого сигнала используется кратковременная АКФ.

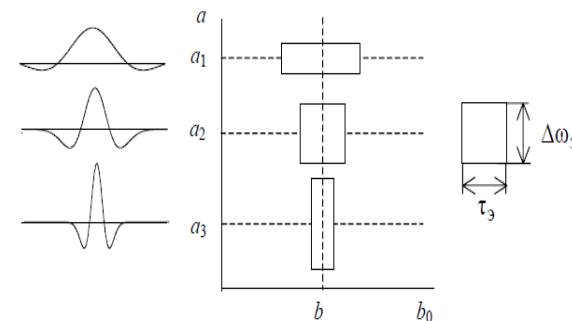
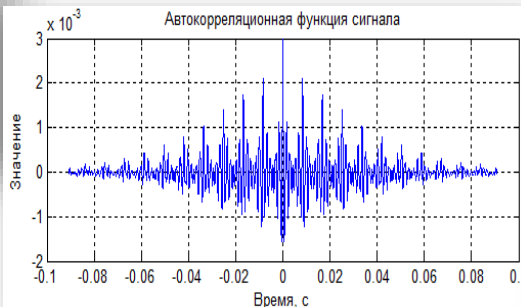
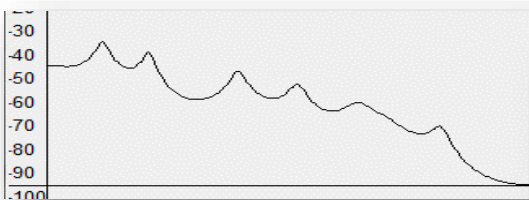
$$R_s(\tau) = \int_{-\infty}^{\infty} s(t) s(t - \tau) dt ,$$

Где  $\tau$  – временной сдвиг.

Базирован на применении непрерывного или дискретного вейвлет-преобразования

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \Psi^* \left( \frac{t-b}{a} \right) dt,$$

где  $s(t)$  – массив временных значений речевого сигнала;  $\Psi$  – базисная функция;  $t$  – время;  $a$  – масштаб;  $b$  – сдвиг;  $a, b \in R, a > 0; s(t) \in L^2(R)$ .



Достоинства

- делает спектры сигнала более плавными и четкими
- даёт удовлетворительные оценки, пока спектр выравнивается достаточно хорошо.

Недостатки

- сложность оценки параметров, использующей линейные процедуры обработки сигнала;
- постоянный расчёт спектра речевого сигнала при подстройке частоты повторения в обратном фильтре;
- определения степени равномерности спектра после обратной фильтрации;
- при частотах основного тона выше 200 Гц данный метод приводит к недостоверным результатам.

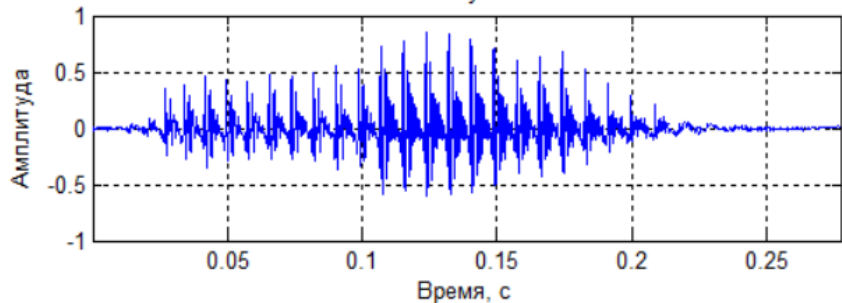
- для вокализованных фонем АКФ имеет четкий максимум в районе задержек, равный периоду ОТ;
- наиболее точно определяет частоту ОТ произнесенных фонем;
- выделить частоту ОТ даже на фоне мощных шумов и различных видах АО

- имеет вычислительную сложность

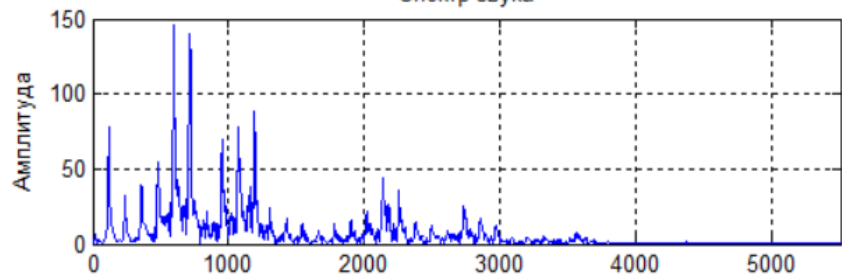
- относительная погрешность оценки частоты основного тона не превышает **0.38%**

- необходимость корректировки окна преобразовании под каждую оцениваемую частоту;
- сложность алгоритма реализации метода;
- большие вычислительные затраты.

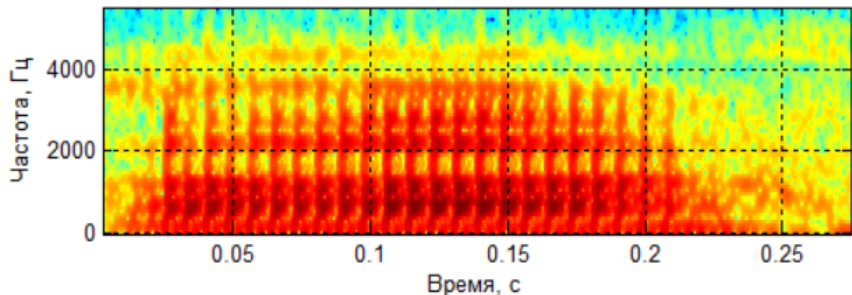
Звук



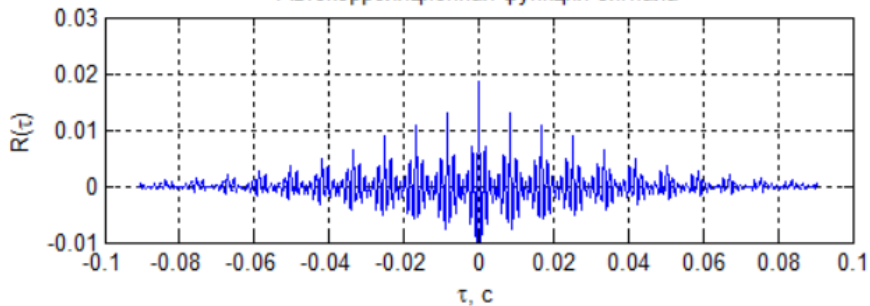
Спектр звука



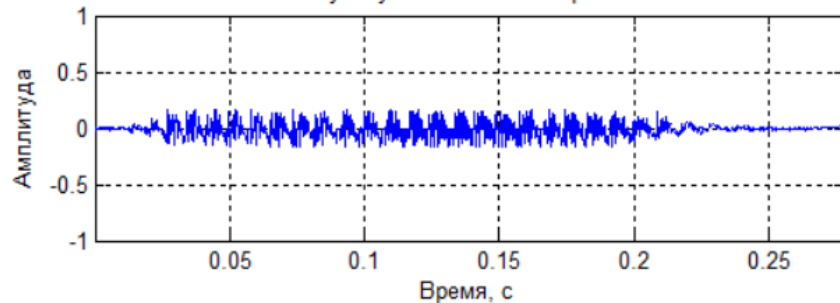
Спектрограмма сигнала



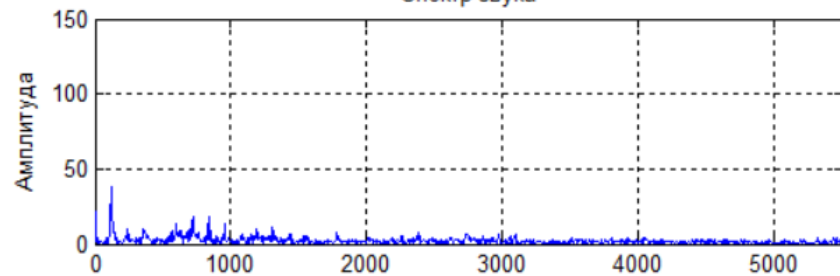
Автокорреляционная функция сигнала



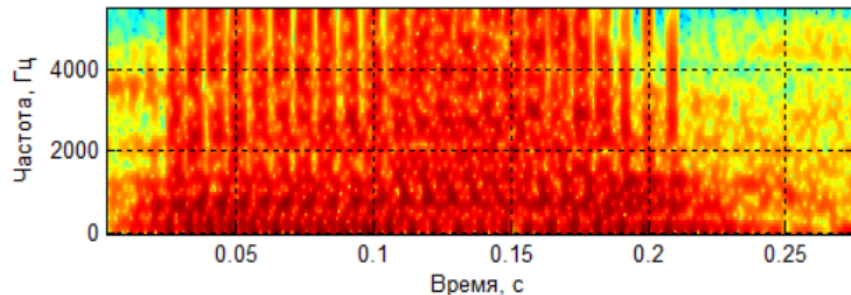
Звук с установленным порогом



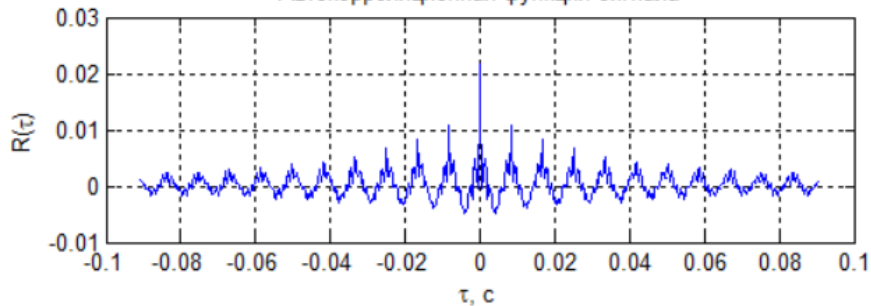
Спектр звука

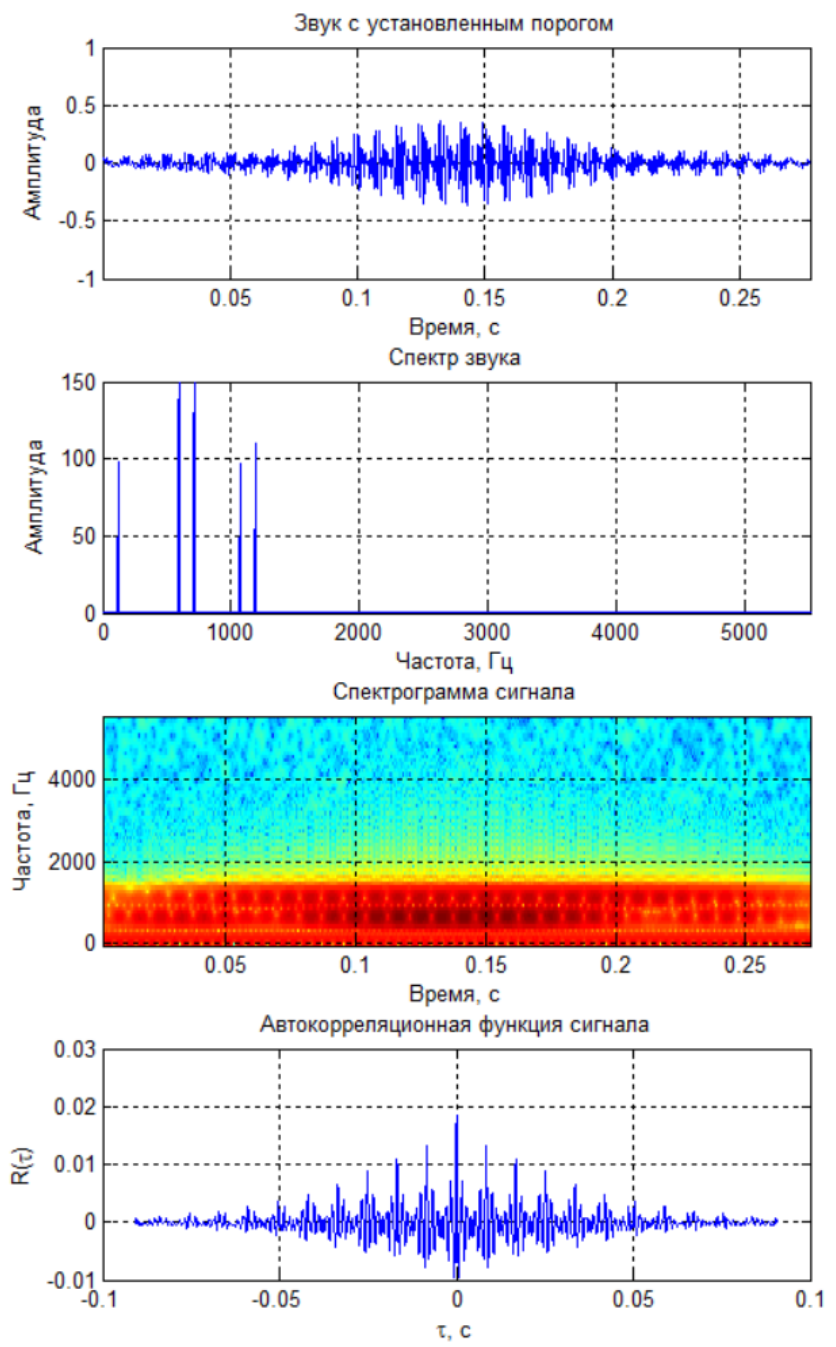
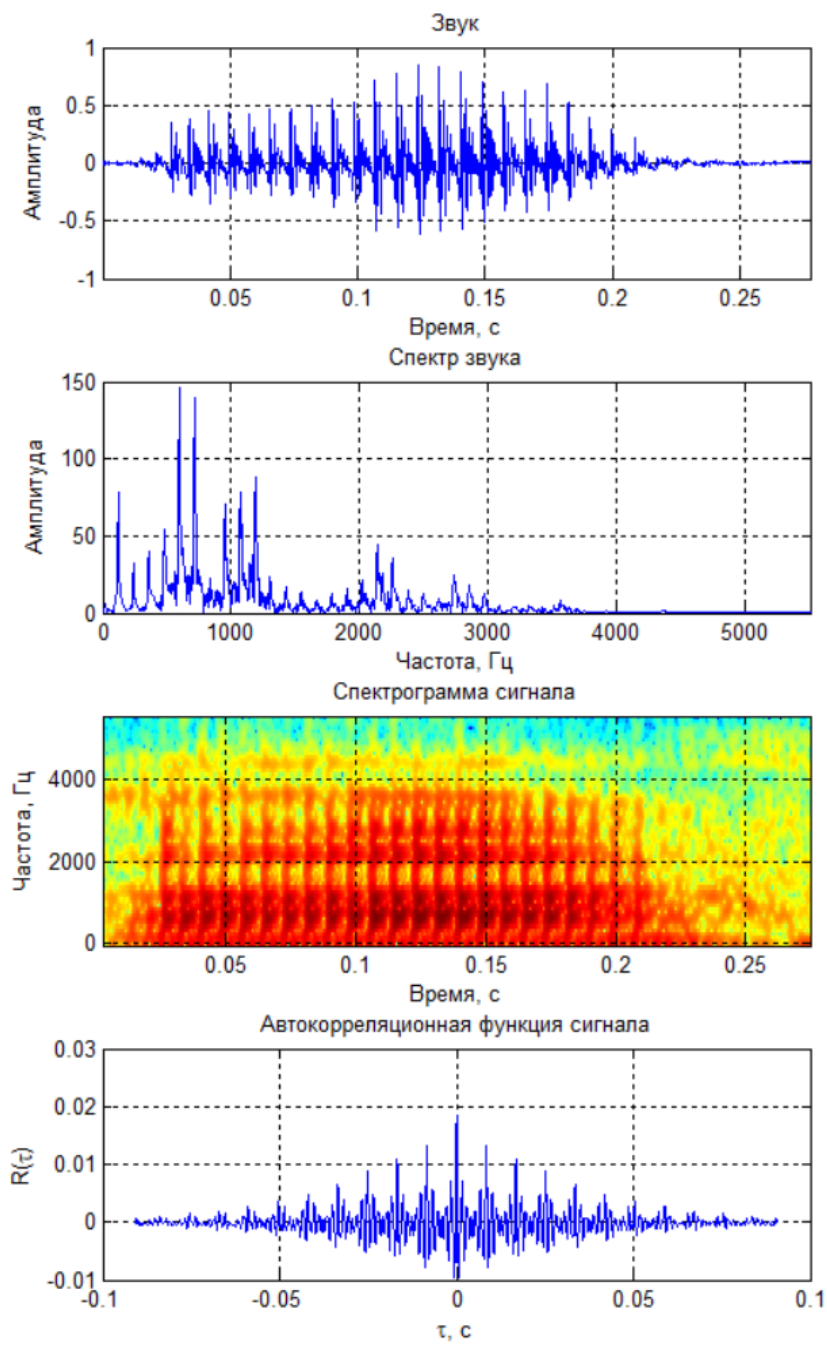


Спектрограмма сигнала



Автокорреляционная функция сигнала





# ПОВЫШЕНИЕ ТОЧНОСТИ ОЦЕНКИ ЗНАЧЕНИЙ ЧАСТОТЫ ОСНОВНОГО ТОНА ГЛАСНЫХ ЗВУКОВ

Первая производная:

$$\frac{\partial R_v(\tau, 0)}{\partial \tau} = \int_{-\infty}^{\infty} s_1(t) \frac{\partial}{\partial \tau} s_1^*(t - \tau) dt.$$

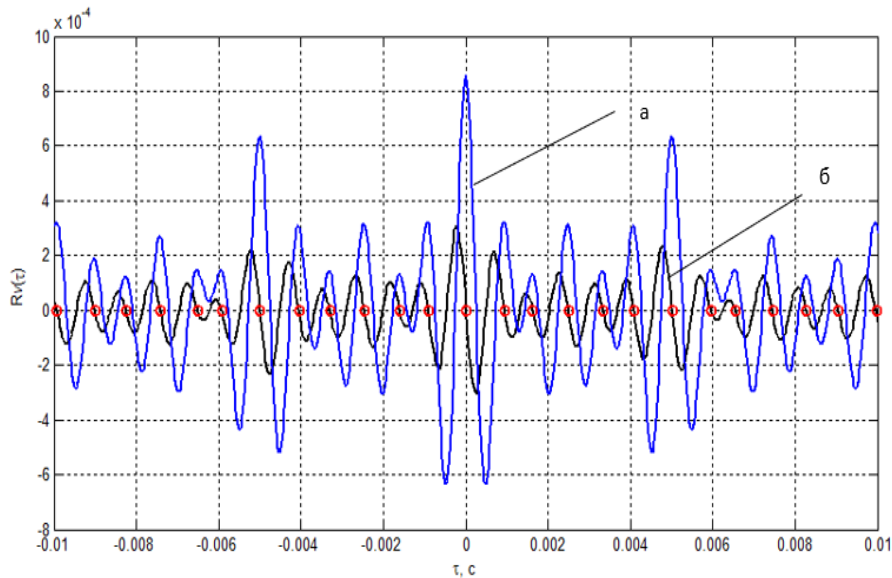


Рис. 1. Совместное отображение АКФ и ее первой производной речевого сигнала:  
а) АКФ; б) первая производная от АКФ

Вторая производная:

$$\frac{\partial^2 R_v(\tau, 0)}{\partial \tau^2} = \int_{-\infty}^{\infty} s_1(t) \frac{\partial^2}{\partial \tau^2} s_1^*(t - \tau) dt.$$

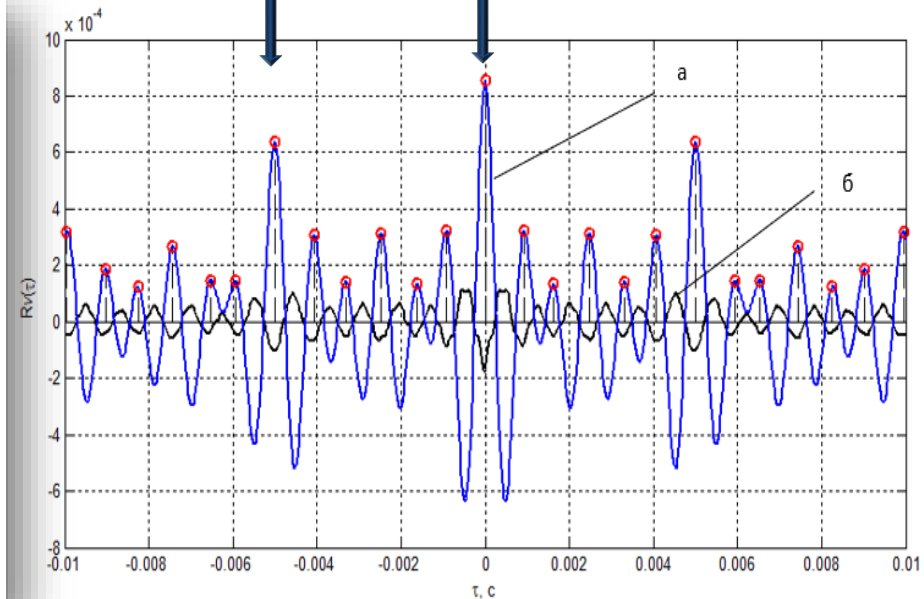


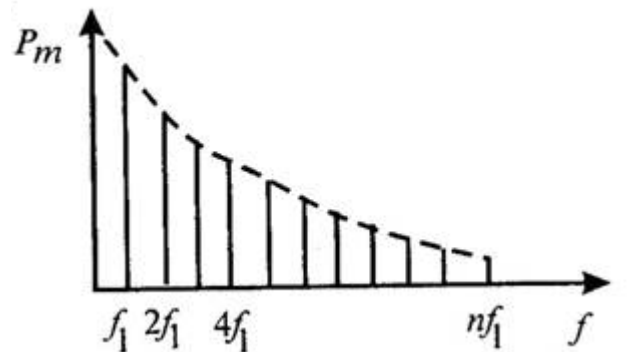
Рис. 2. Совместное отображение АКФ и ее второй производной речевого сигнала:  
а) АКФ; б) вторая производная от АКФ

## Выводы:

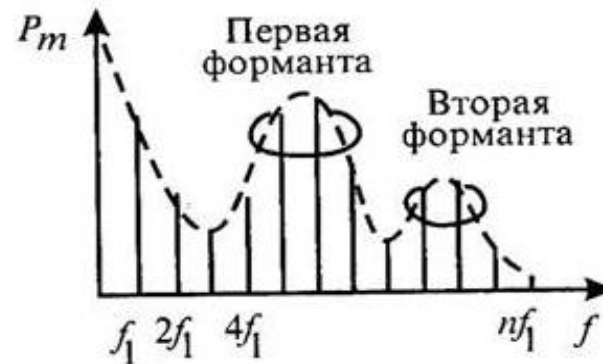
1. Разброс вычисленных значений частоты ОТ (за исключением метода вейвлет-преобразований) составил  $\pm 1,37\%$ .

# ФОРМАНТЫ

**Форманты** – области максимальной концентрации энергии в ограниченной частотной области, которые характеризуются *амплитудой  $A_i$* , *частотой  $F_i$*  и *шириной полосы  $\Delta F_i$* .



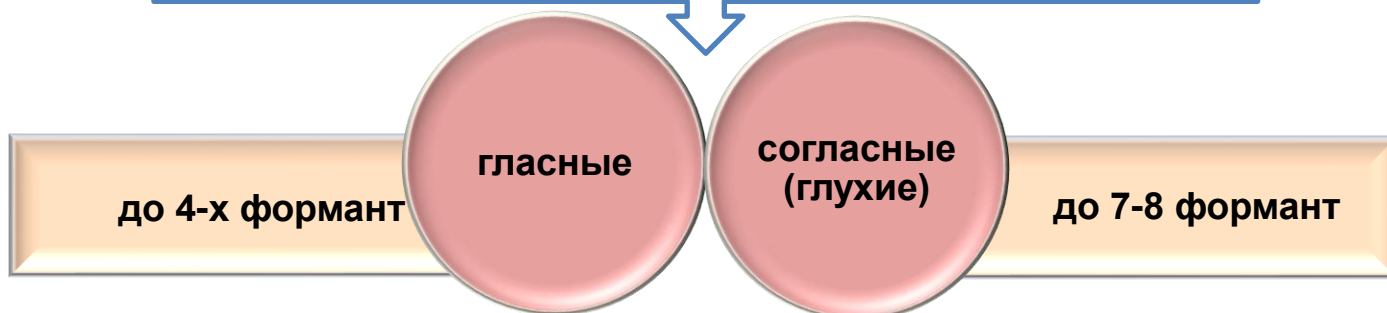
Формируемый голосовыми связками



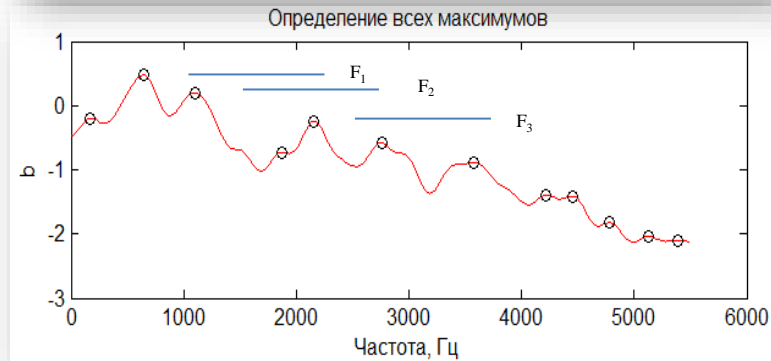
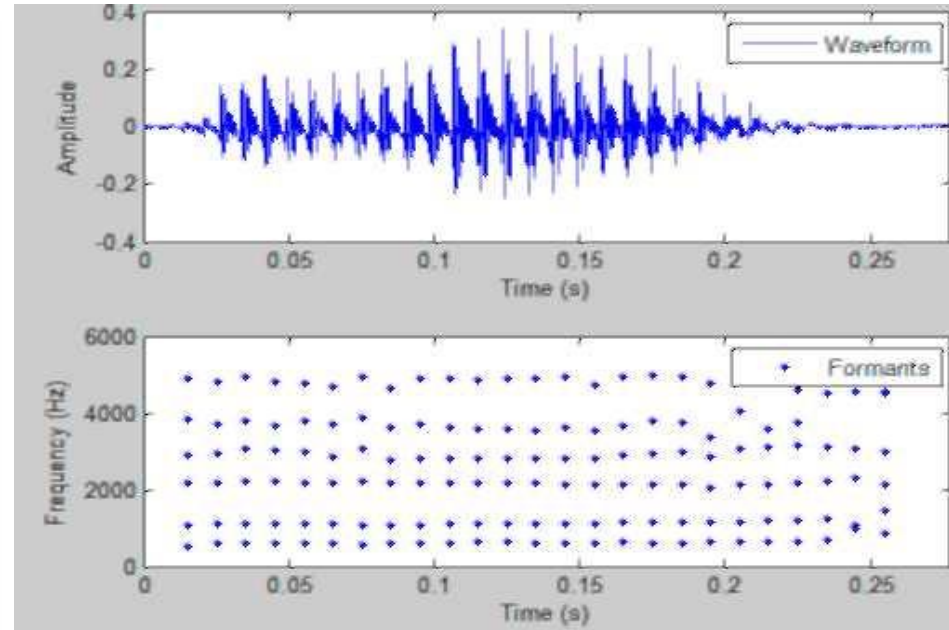
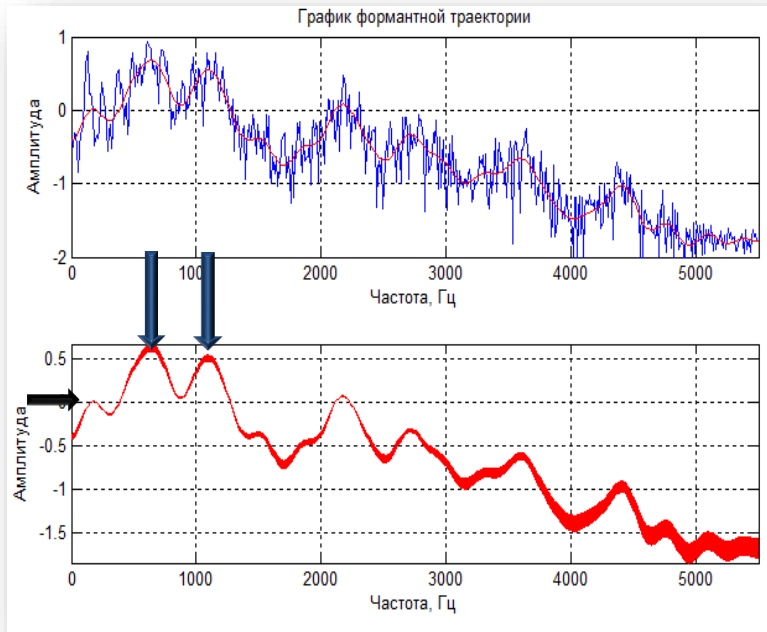
Излучаемый ртом

Рис 1. Спектр звука

Различные звуки имеют разное число формант



# ОПРЕДЕЛЕНИЕ ЗНАЧЕНИЙ ЧАСТОТЫ ОСНОВНОГО ТОНА НА ПРИМЕРЕ ГЛАСНЫХ ЗВУКОВ



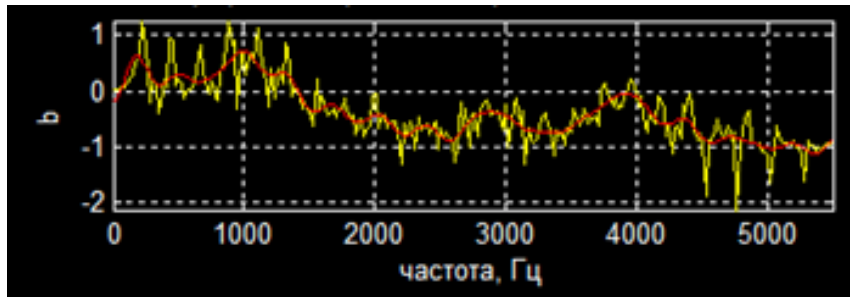
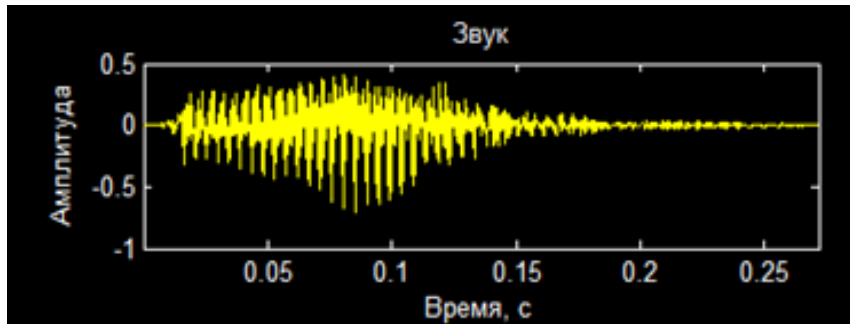
Форманты	Для звонких	Для глухих
<b>F1</b>	150-700 Гц	150-700 Гц
<b>F2</b>	700-1400 Гц	700-1400 Гц
<b>F3</b>	1400-2300 Гц	1400-3000 Гц
<b>F4</b>	2300-4000 Гц	3000-6000 Гц

## Выводы:

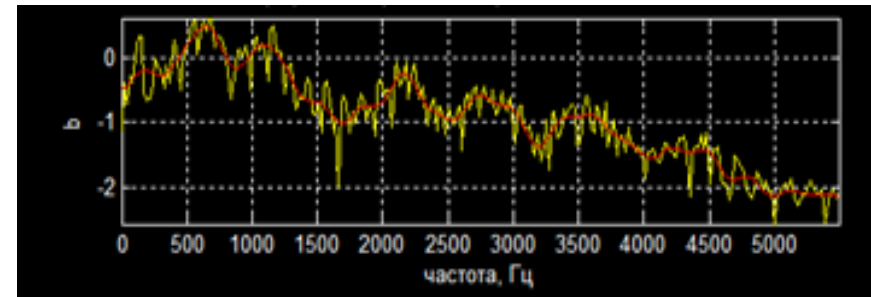
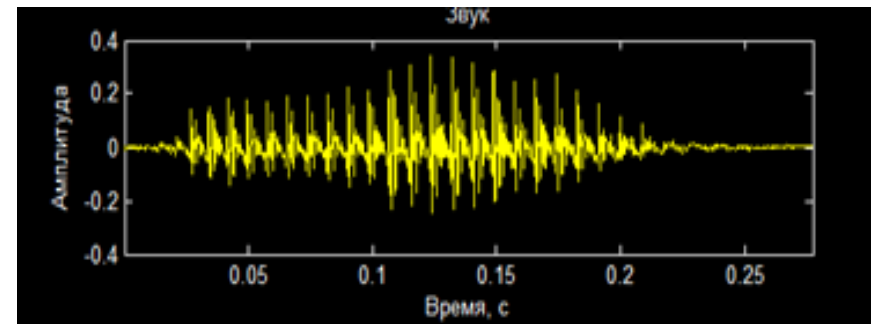
1. Диапазон для определения основных формант для звонких звуков **100-3000 Гц**.
2. Ограничение диапазонов для определения основных формант позволяет снизить уровень шумов.

# ПРИМЕР ФОРМАНТНЫХ ТРАЕКТОРИЙ

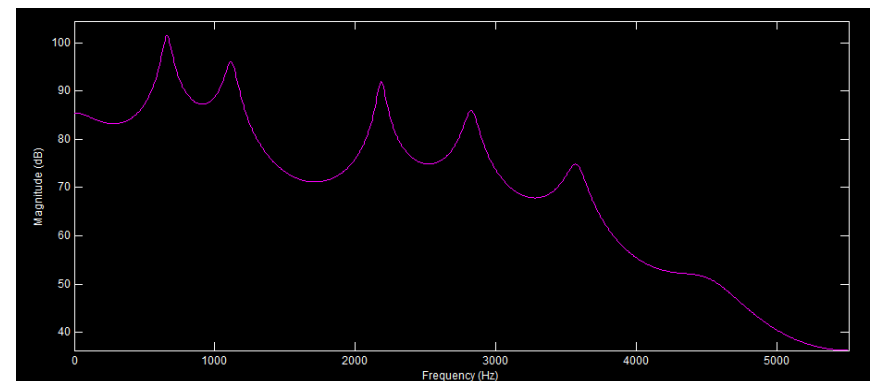
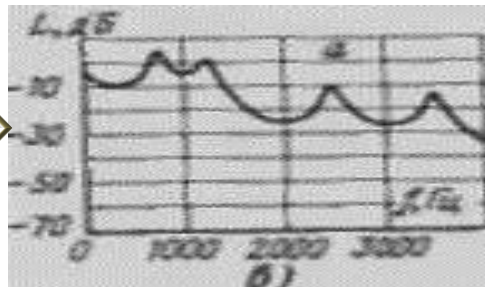
Женский голос (звук а)



Мужской голос (звук а)

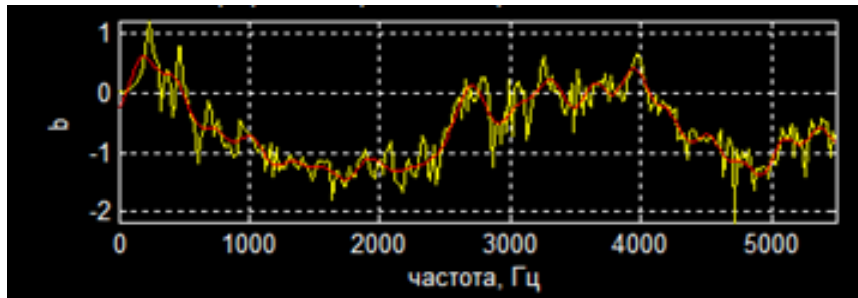
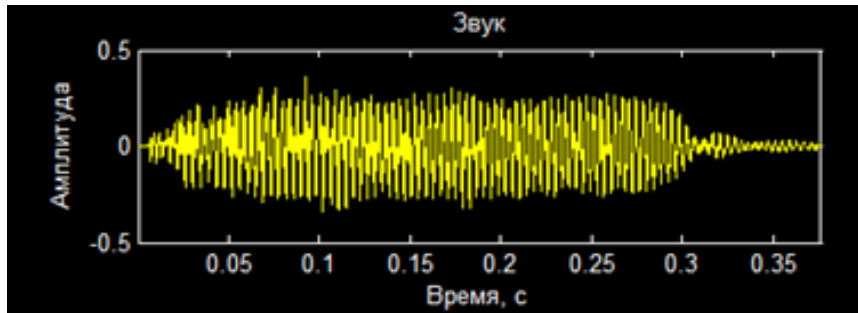


Сапожков М. А

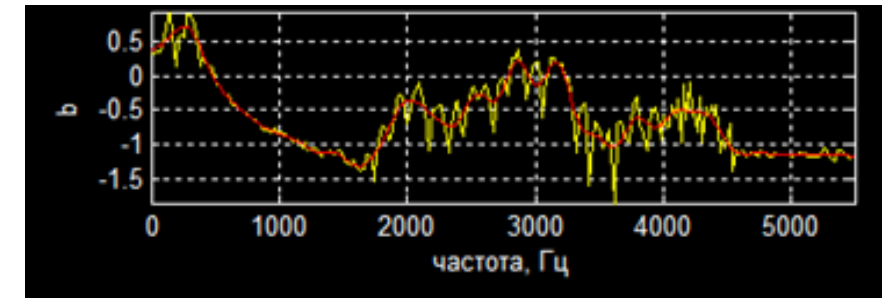
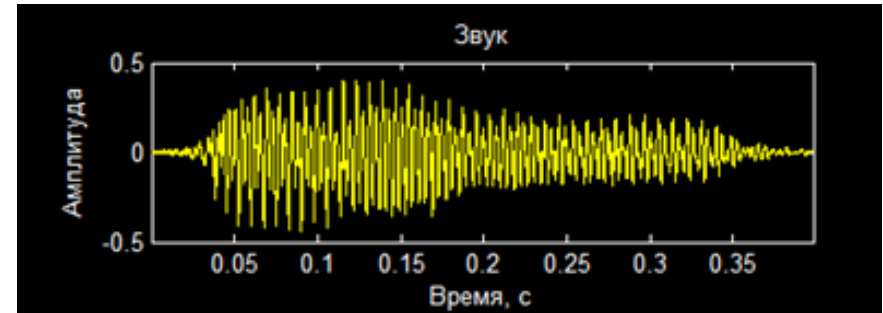


# ПРИМЕР ФОРМАНТНЫХ ТРАЕКТОРИЙ

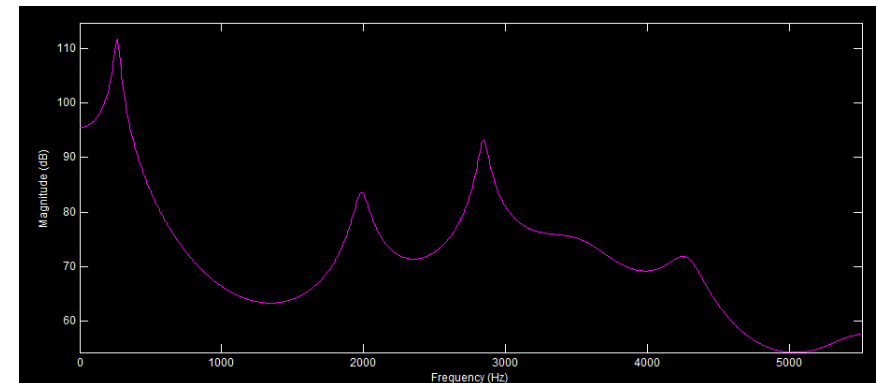
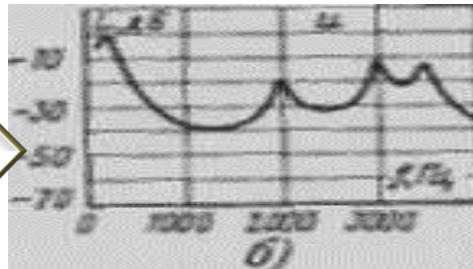
Женский голос (звук и)



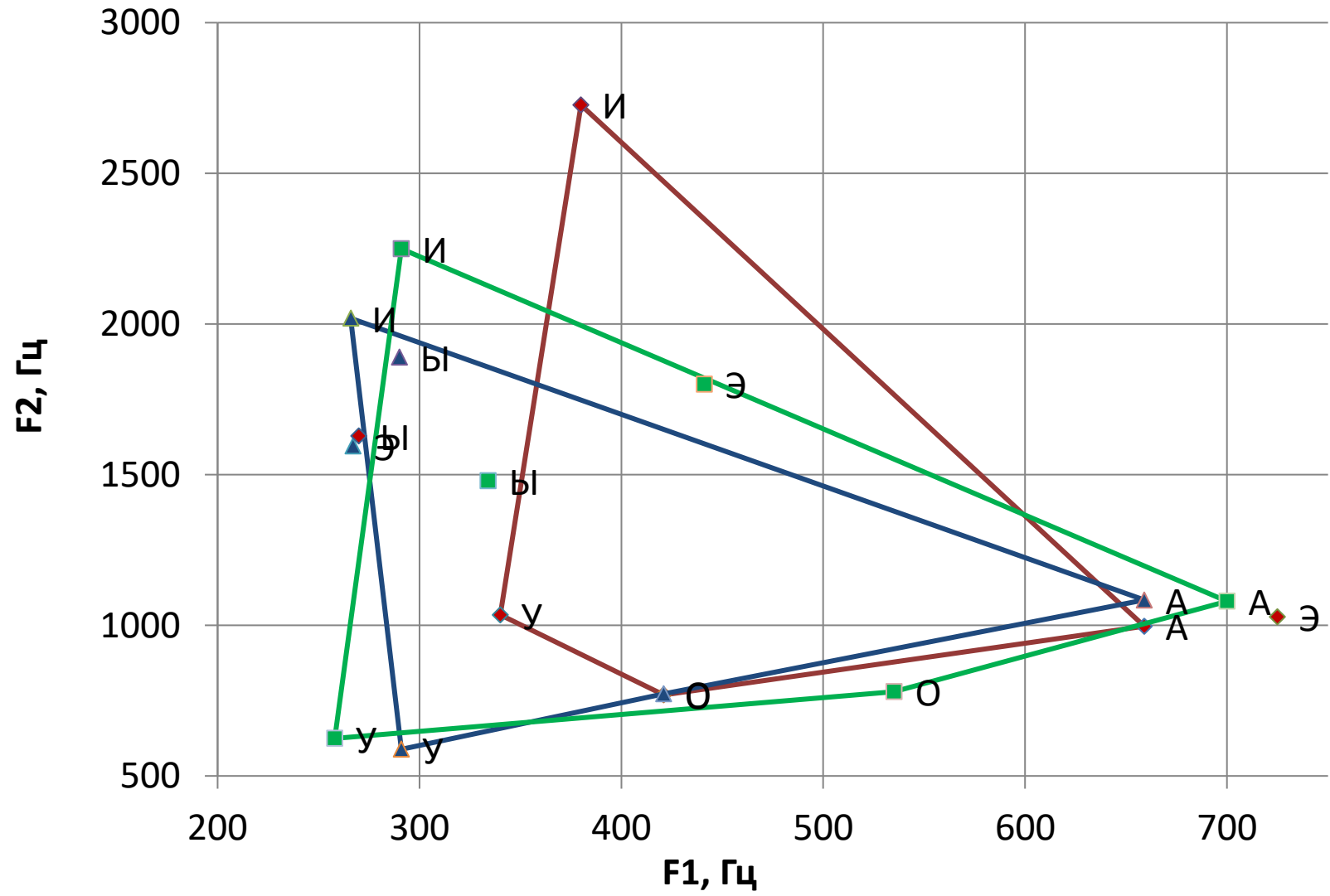
Мужской голос (звук и)



Сапожков М.А.



# ФОРМАНТНЫЕ «ЧЕТЫРЕХУГОЛЬНИКИ»



# ПОНЯТИЕ ВЕЙВЛЕТ-АНАЛИЗА

В переводе с английского *Wavelet* – всплеск, поэтому в отечественной литературе встречается термин «**разложение по всплескам**» наряду с **вейвлет-анализом**.

## Посылка:

Использование такого **базиса** который позволит осуществить **переход от частотного анализа к масштабному**, т.е. некоторой “стандартной” математической **функции  $\psi(x)$ , изменяемой по масштабу и сдвигу на некоторую величину**.

В **1980 году Гроссман и Марлет** определили такие функции как *Wavelet*-функции. В конце 80-х г.г. **Мейер и Добеши** на основе исследований Марлета создали ортогональные базисы *Wavelet*-функций, которые и стали **основой современных *Wavelet*-функций**.

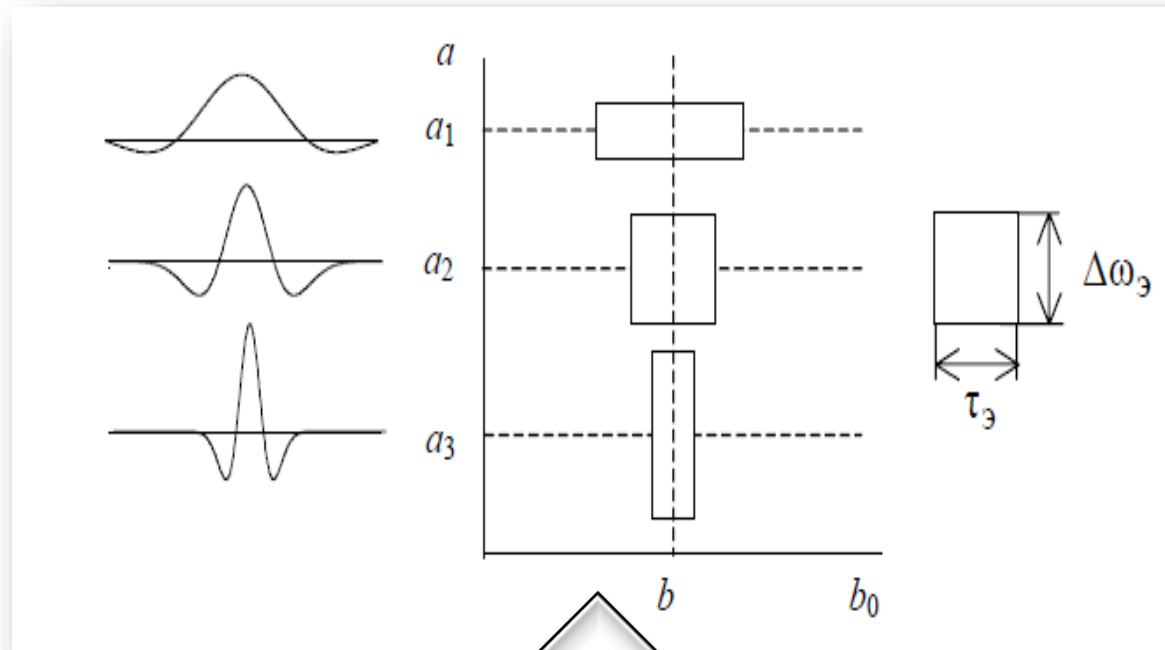
В период **60-х–80-х годов Вейс и Кофман** исследовали простейшие элементы функционального пространства, названные ими атомами, с целью обнаружить атомы для произвольной функции и найти “правила сборки”, позволяющие реконструировать все элементы функционального пространства, используя эти атомы.

**Пол Леви**, используя масштабируемую базисную функцию типа функции Хаара, исследовал разновидность случайного сигнала – броуновское движение. Он **обнаружил преимущество в применении базисных функций Хаара перед функциями Фурье**.

**Первое упоминание** об этих функциях появилось в работах **Хаара в 1909 году**.

В 30-е годы начались более детальные исследования возможностей представления сигналов с использованием базисных масштабируемых функций.

# ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ



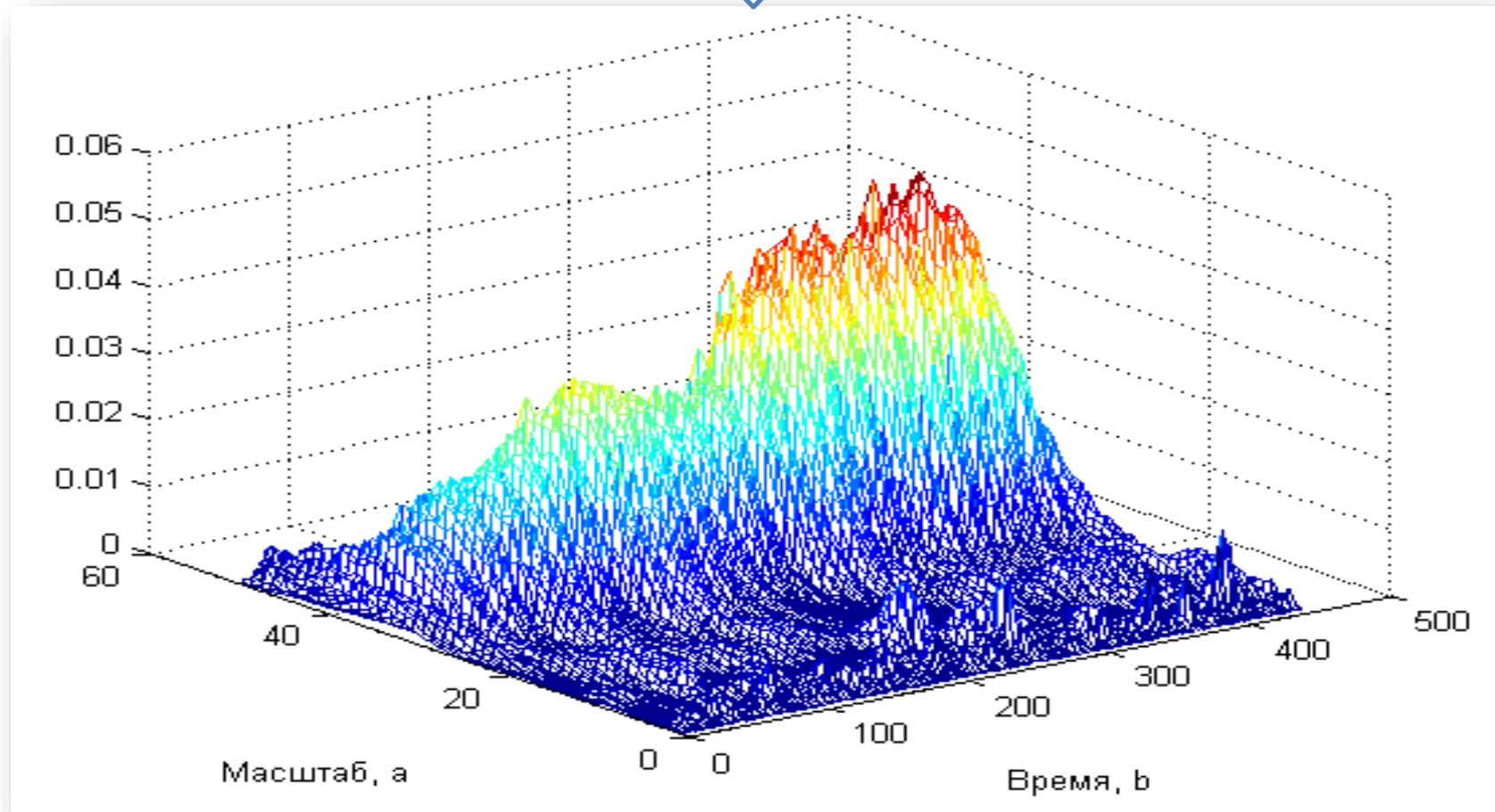
Непрерывное вейвлет-преобразование:

$$f(t) \in L^2(\mathbb{R}) \quad f(t) \in L^2(\mathbb{R})$$

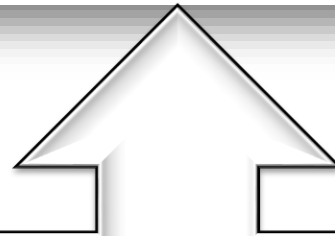
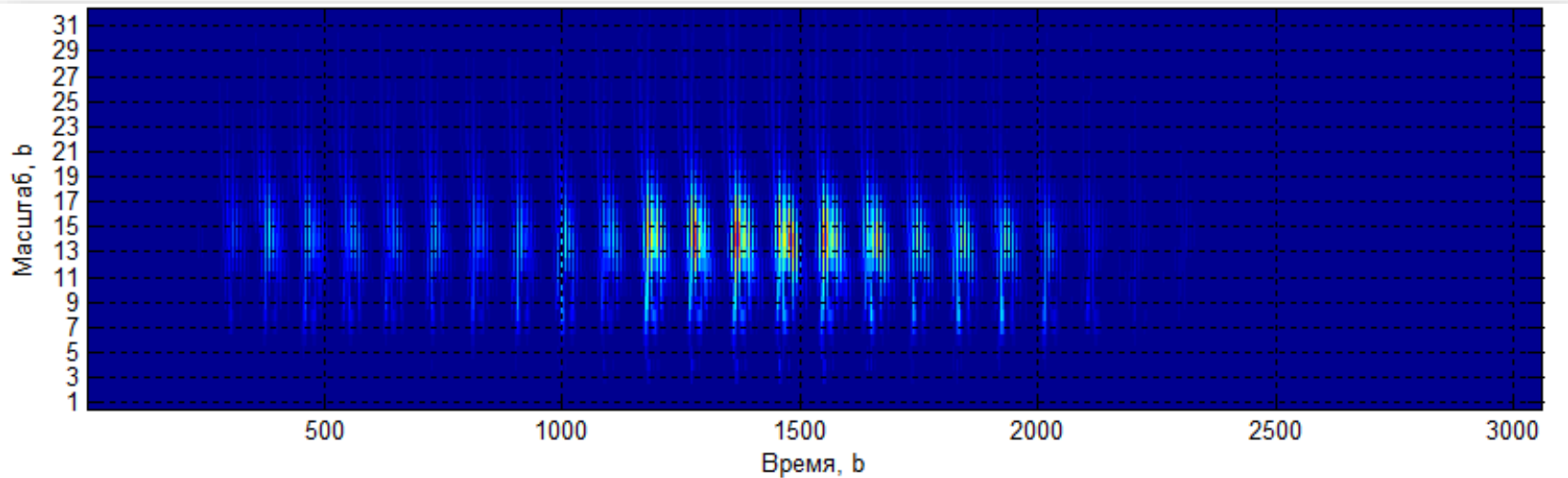
где  $a$  – масштаб,  $b$  – сдвиг  $W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi^*\left(\frac{t-b}{a}\right) dt$   
 $s(t)$  – исходный сигнал,  $\psi$  – базисная функция,  $t$  – время.

# ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ ФОНЕМЫ «А»

Искаженная объемная скейлограмма, представляющая фрагмент фонемы «а» русского языка в виде набора мультифрактальных структур



# ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТОВ В ЗАДАЧАХ ОБРАБОТКИ И РАСПОЗНАВАНИЯ РЕЧИ

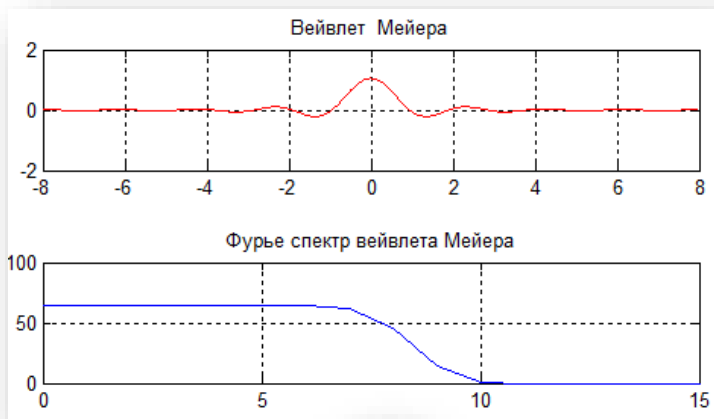


Дискретное вейвлет-преобразование:  $W_A(a, b) = \frac{1}{n(a, b)} \sum_{k=0}^{N-1} s_k \psi^* \left( \frac{t_k - b}{a} \right)$ ,

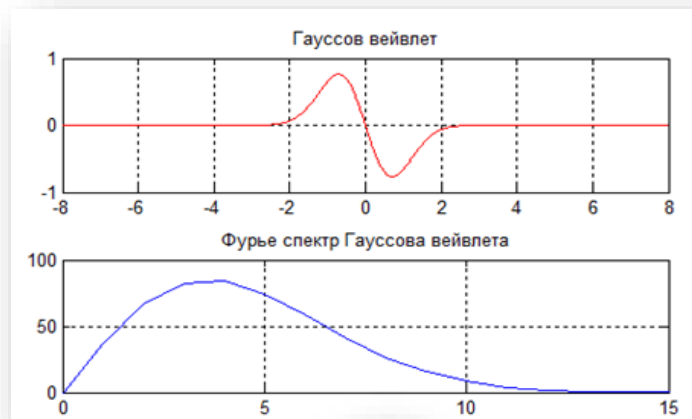
где  $n(a, b) = \sum_{k=0}^{N-1} e^{-\frac{1}{B} \left( \frac{t_k - b}{a} \right)^2}$ ,  $\psi(t)$  – вейвлет,

$s_k = s(t_k)$ ,  $t_k = \Delta t k$ , – фрагмент гласного звука представленный временным рядом со значениями функции, следующими друг за другом с постоянным интервалом времени  $\Delta t$ .  
 $k = 0, 1, \dots, N - 1$

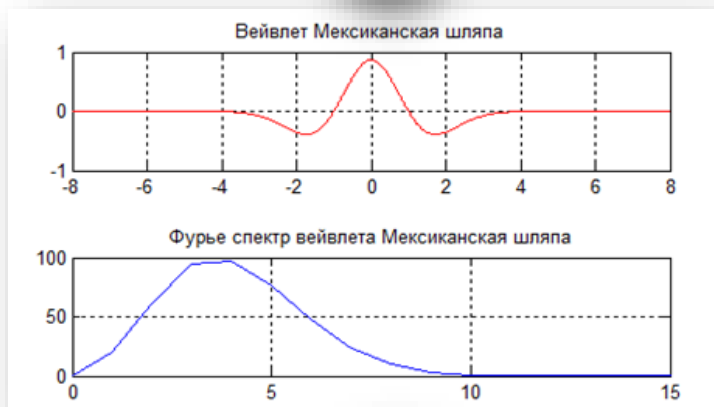
# ПРИМЕРЫ ВЕЙВЛЕТ-ФУНКЦИЙ



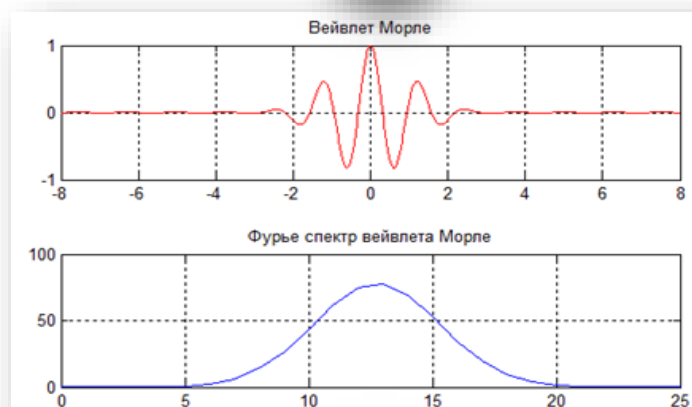
а)



б)



в)

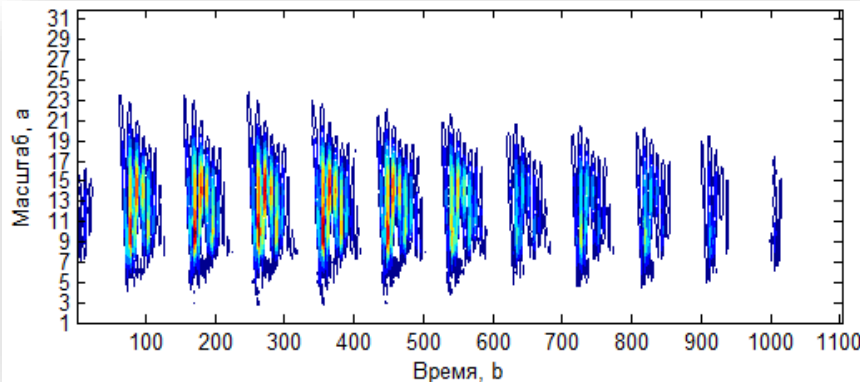


г)

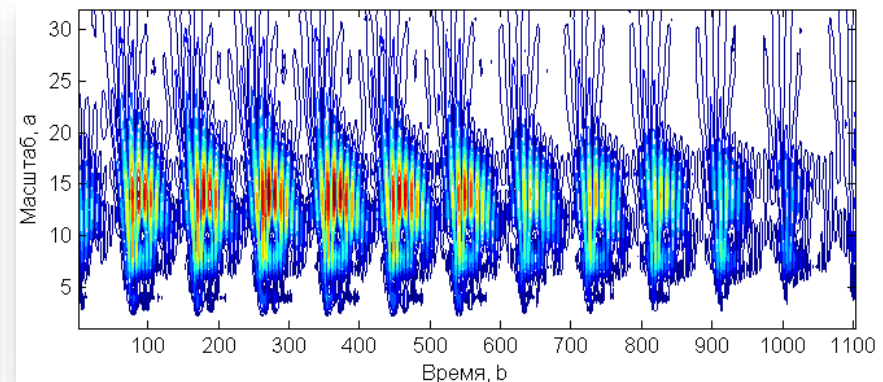
Исследованы вейвлеты: Меера, Гауссова типа, «Мексиканской шляпы» и Морле.

# ИНФОРМАТИВНОСТЬ ВЕЙВЛЕТ-ПРЕДСТАВЛЕНИЯ ПРИ ИЗМЕНЕНИИ МАСШТАБНЫХ КОЭФФИЦИЕНТОВ А И В

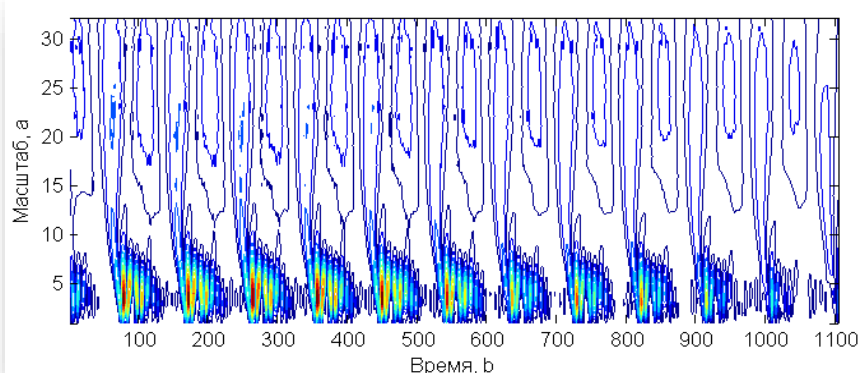
Примеры скалограмм звука «а», полученных с использованием различных материнских вейвлетов при одних и тех же масштабных коэффициентах а и b



а) с использованием вейвлет-преобразования на основе вейвлета **Мейера**



б) с использованием вейвлет-преобразования на основе комплексного вейвлета **Морле**



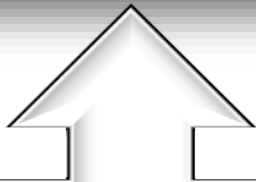
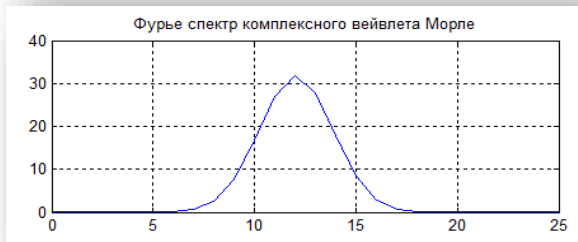
с) с использованием вейвлет-преобразования на основе вейвлета «**Мексиканской шляпы**»

Оценка локального спектра энергии проводится на основании **скалограммы**:

$$S(a_i, b_j) = |W_A(a_i, b_j)|^2,$$

которую можно отобразить в трехмерном пространстве координат  $(a, b, S)$  или в виде топологической карты, изобразив поверхность  $S(a, b)$  в координатах  $(a, b)$ .

# ОСОБЕННОСТИ КОМПЛЕКСНОГО ВЕЙВЛЕТА МОРЛЕ



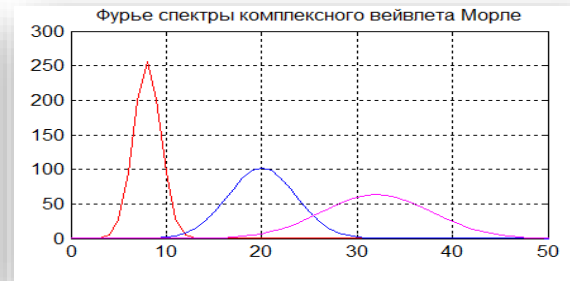
Комплексный вейвлет Морле описывается функцией:

$$\psi(t) = \frac{1}{\sqrt{\pi B}} e^{-\frac{t^2}{B}} e^{i2\pi f_0 t}$$

где  $B$  — ширина полосы частот;  
 $f_0$  — центральная частота вейвлета  
 $t$  — время.

## Преимущества:

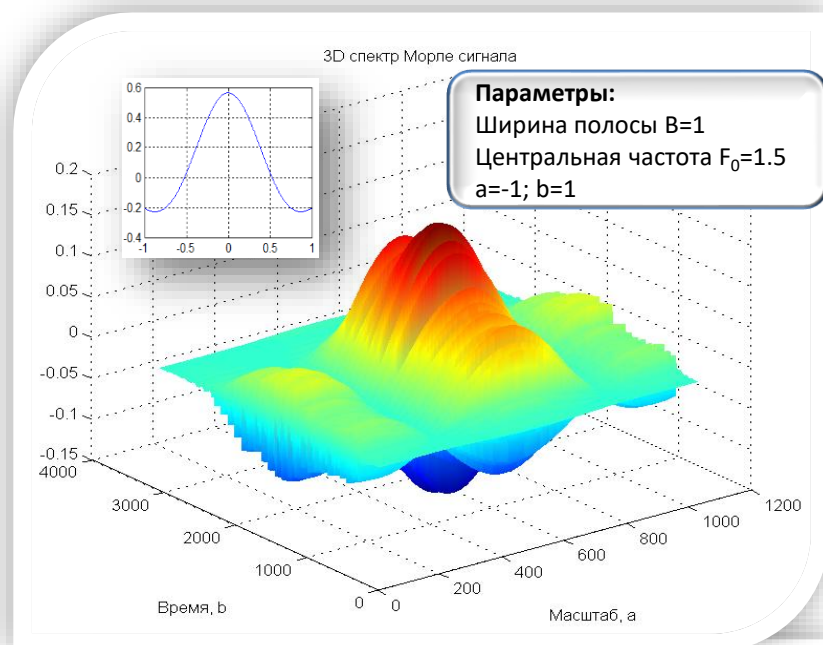
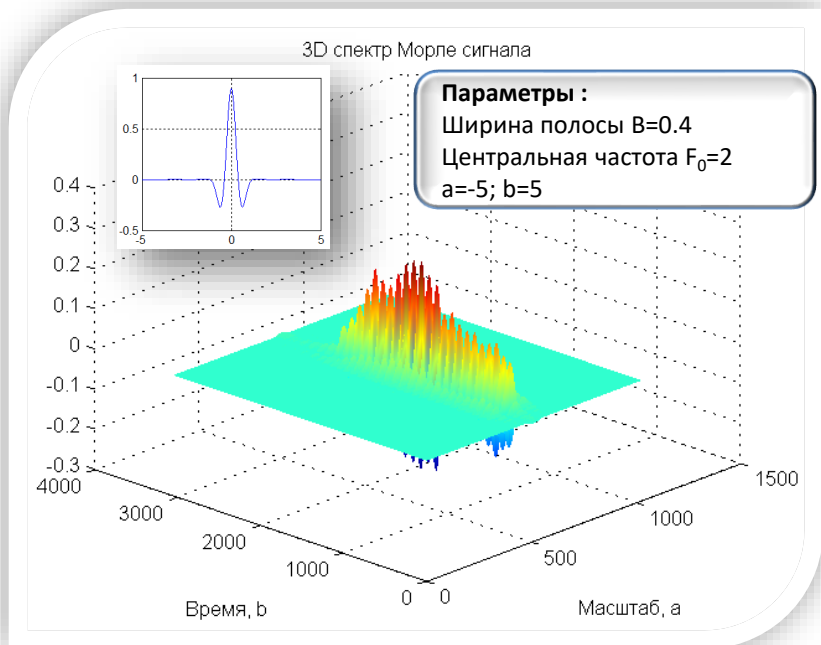
- имеет более узкий Фурье образ и продолжителен во временной области;
- присутствие доминирующей частоты позволяет варьировать избирательностью вейвлета в частотной области;
- обладает частотной локализацией лучшей среди других базисов;
- на практике при  $f_0 \gg 0$  вейвлет Морле может быть использован с минимальной погрешностью;
- позволяет разделить фазовые и амплитудные компоненты сигнала при выполнении вейвлет-преобразования;
- имеет близкое сходство с речевыми фрагментами.



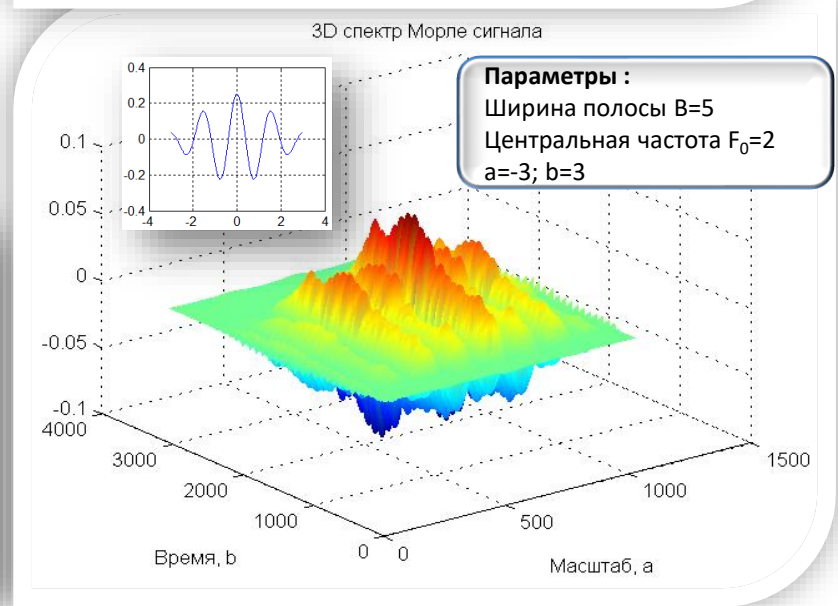
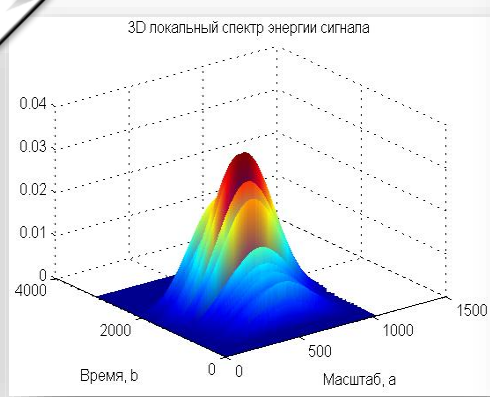
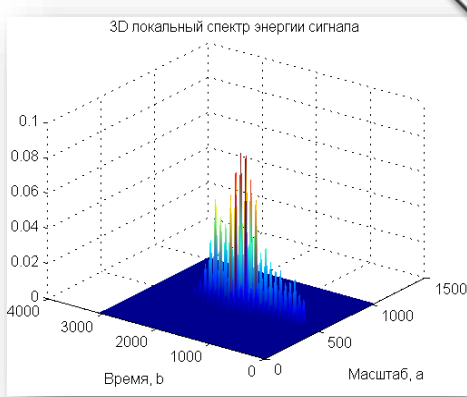
## Вывод:

Учитывая все перечисленные преимущества, для оценки тонкой структуры речевых сигналов в условиях воздействующих факторов рекомендовано использовать **комплексный вейвлет Морле**.

# ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ ФОНЕМЫ «А» ПРИ ИЗМЕНЕНИИ ПАРАМЕТРОВ КОМПЛЕКСНОГО ВЕЙВЛЕТА МОРЛЕ



Оценка локального спектра энергии проводится на основании скалограммы:  $S(a_i, b_j) = |W_A(a_i, b_j)|^2$



# ПОЛУЧЕНИЕ ДОПОЛНИТЕЛЬНЫХ СВЕДЕНИЙ О СИГНАЛЕ НА ОСНОВАНИИ ЕГО СКЕЙЛИНГОВЫХ ПАРАМЕТРОВ

Двумерные срезы трехмерной скейлограммы заданного фрагмента звука «а» при различных параметрах вейвлета Морле.

Используя скалограмму, можно проводить оценку глобального спектра энергии и построить скейлограмму:

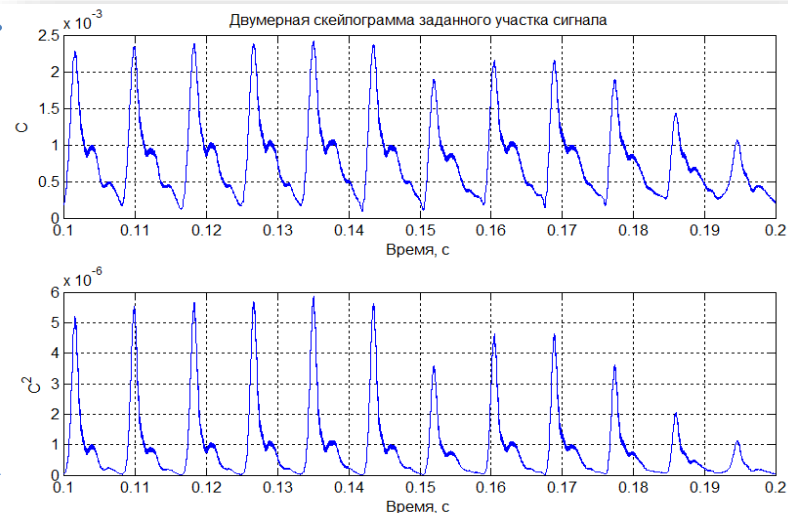
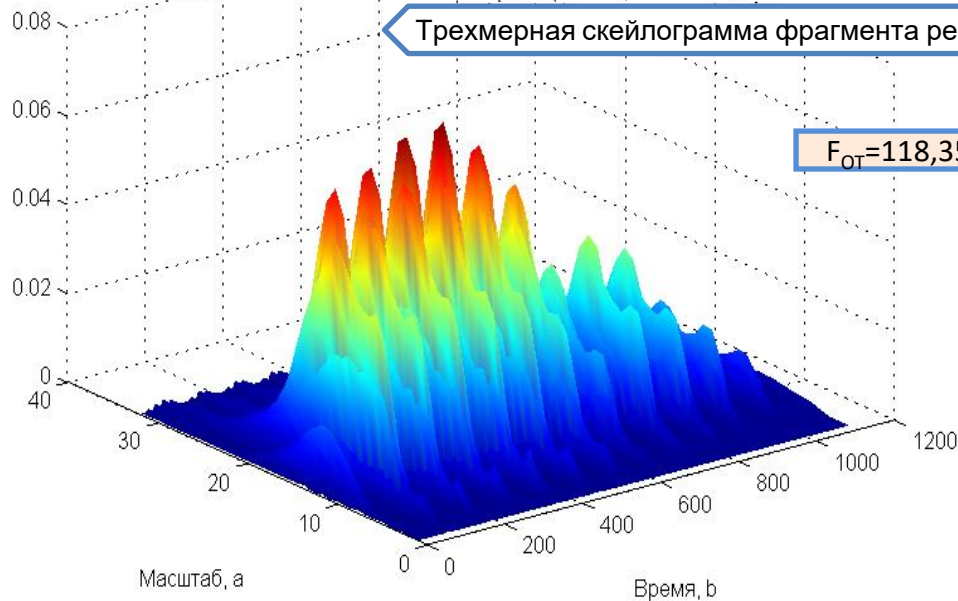
$$G(a_i) = \frac{1}{N^*} \sum_j S(a_i, b_j),$$

где  $N^*$  – число точек, по которому осуществляется осреднение.

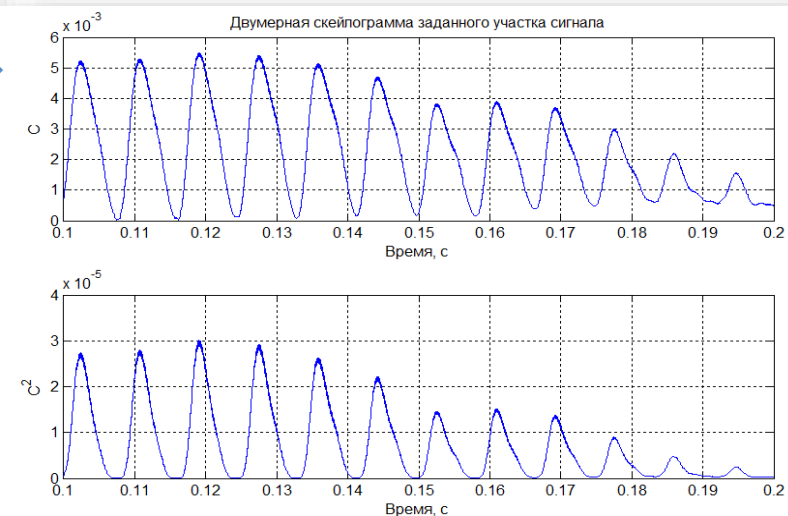
$F_{OT}=118,55$

Трехмерная скейлограмма фрагмента речи

$F_{OT}=118,35$

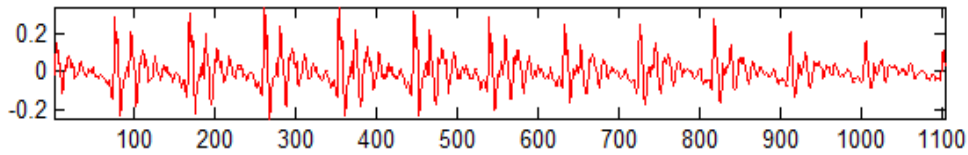


а) параметры вейвлета:  $B=1$ ;  $f_0=1.5$

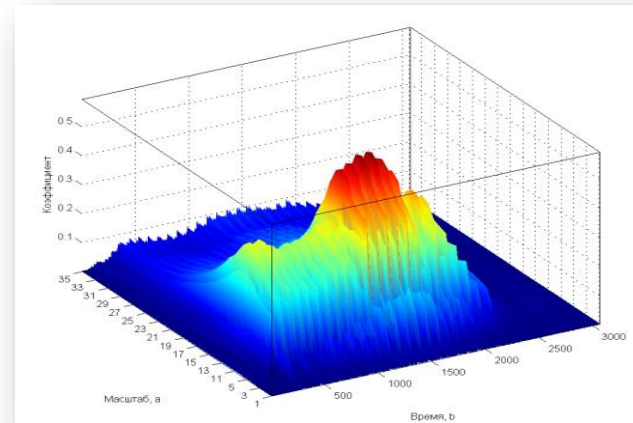
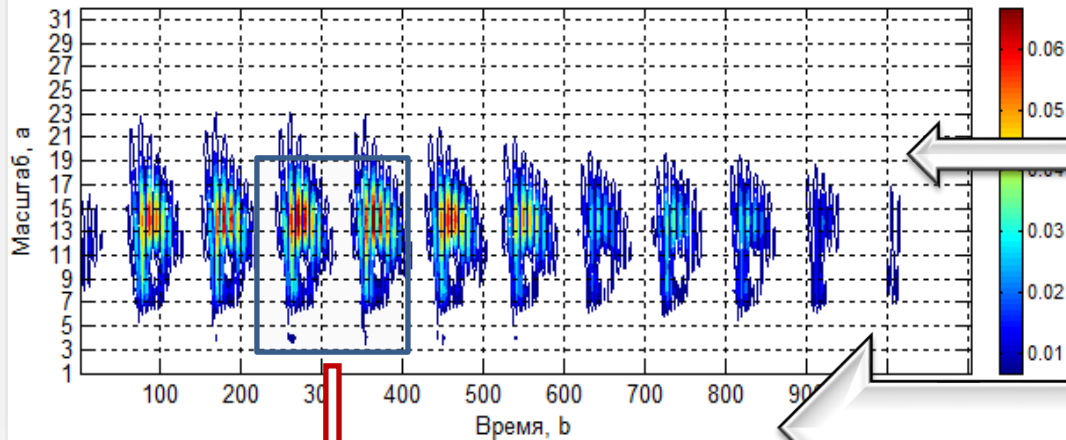


б) параметры вейвлета:  $B=2$ ;  $f_0=0.5$

# ОЦЕНКА ЛОКАЛЬНОГО СПЕКТРА ЭНЕРГИИ ФОНЕМЫ «А»



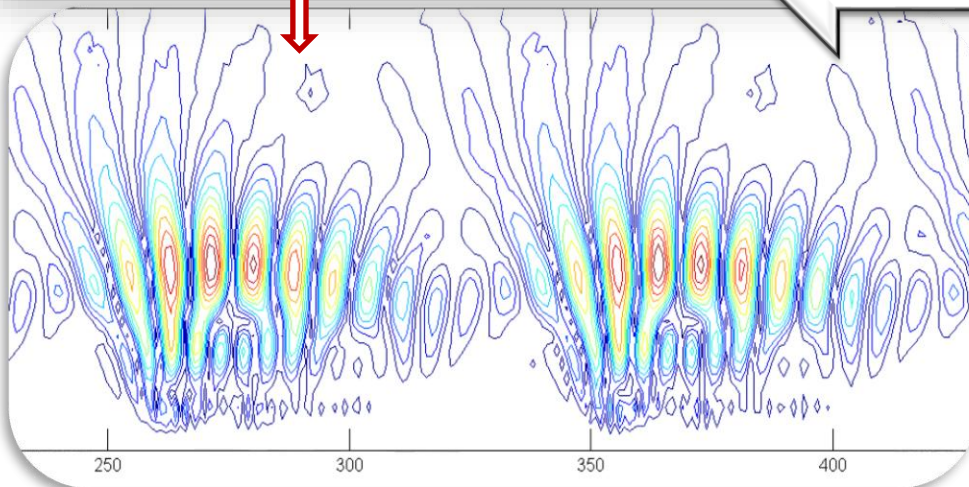
Скалограмма



Пример скалограммы фрагмента  
фонемы «а»

Можно отсечь влияние контуров, выделив  
те точки скалограммы, в которых она  
имеет максимумы по переменным  $a$  и  $b$ ,  
построив **скелетон**:

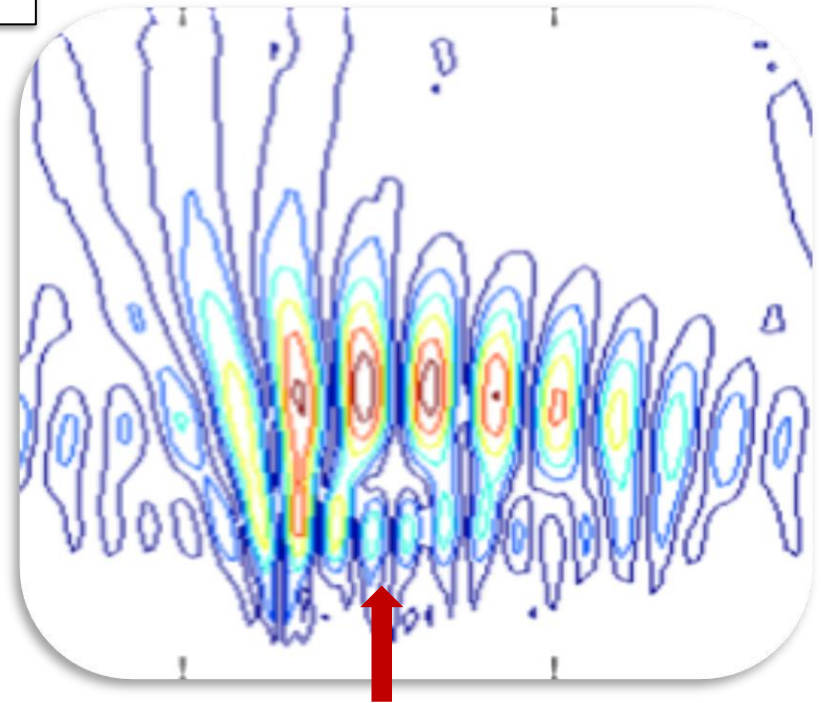
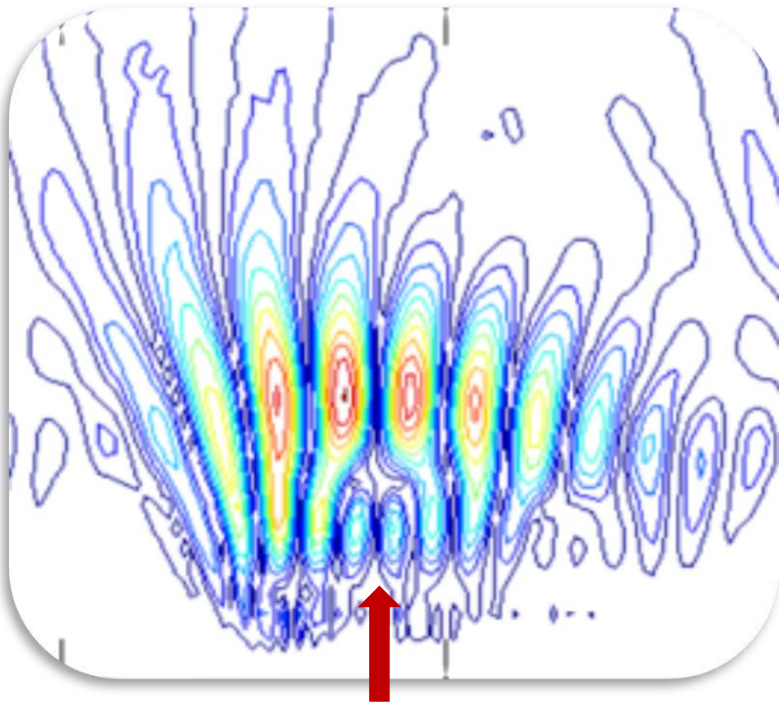
$$S_c(a_i, b_j) = \begin{cases} S_{ij}, & \text{если } S_{i-1,j} < S_{i,j} > S_{i+1,j} \\ & \text{или } S_{i,j-1} < S_{i,j} > S_{i,j-1} \\ 0, & \text{в остальных случаях.} \end{cases}$$



# ИДЕНТИФИКАЦИЯ ПО ГОЛОСУ НА ПРИМЕРЕ ФОНЕМЫ «А» МУЖСКОЙ ГОЛОС

Фонема «а», произнесенная одним человеком в разных условиях

$F_{OT}=118$  Гц

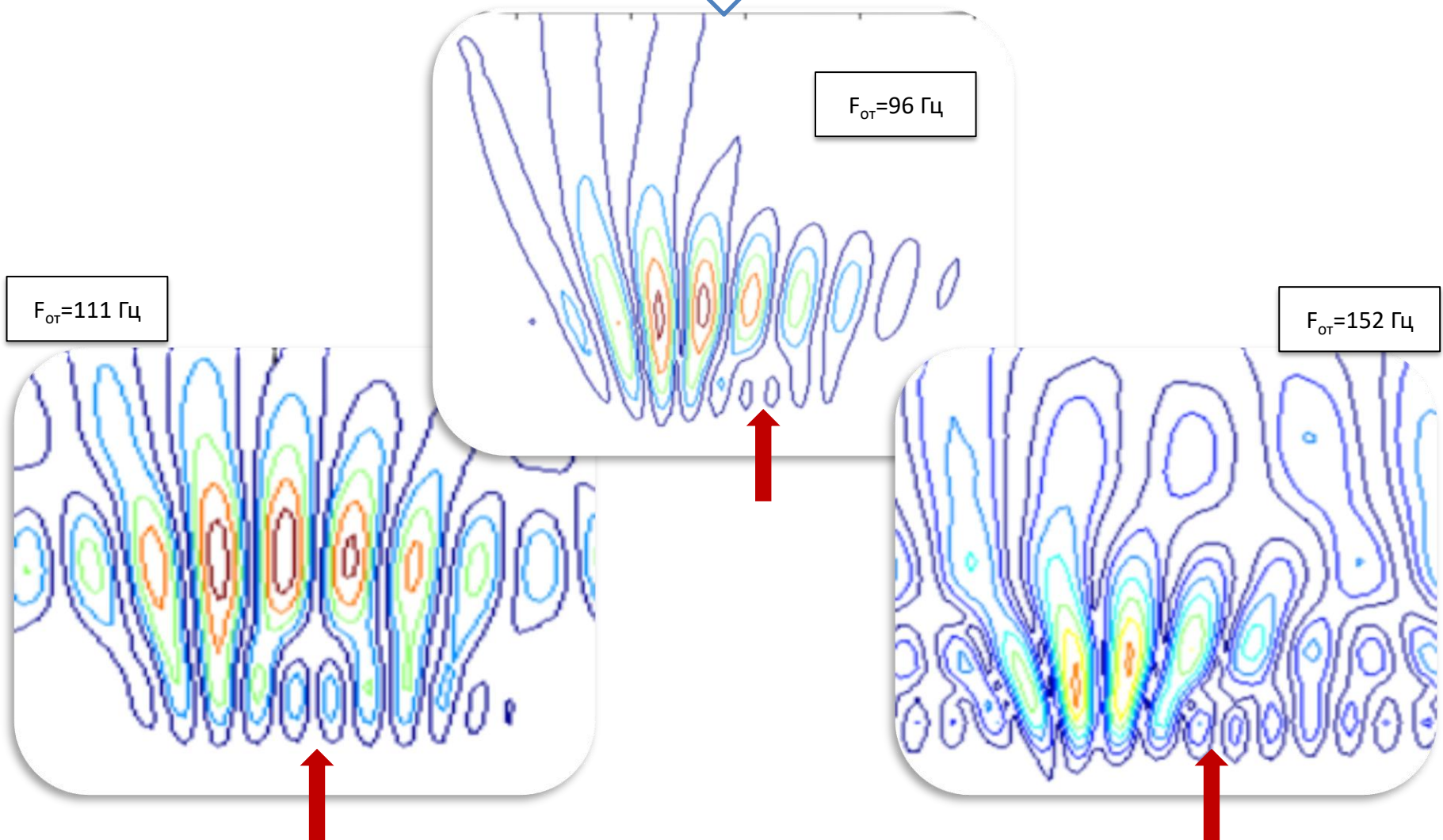


## Вывод:

Применение **вейвлет-преобразования** позволяет проводить **оценку скейлинговых** параметров сигналов и с высокой точностью определять значение частоты OT – относительная погрешность оценки не превышает **0,38%**.

# ИДЕНТИФИКАЦИЯ ПО ГОЛОСУ НА ПРИМЕРЕ ФОНЕМЫ «А» МУЖСКОЙ ГОЛОС

Фонема «а», произнесенная разными людьми



## ВЫВОДЫ

**Разброс вычисленных значений частоты ОТ** рассмотренными методами (за исключением метода вейвлет-преобразований) составил  $\pm 1,37\%$ .

На основании обработки экспериментальных данных, **наилучшие результаты оценки ОТ в условиях шумов получены методом АКФ**, позволяющим оценить периодичность сигнала в зависимости от его задержки. Данный метод **не требует больших вычислительных ресурсов** позволяет наиболее **точно определить частоту ОТ** (относительная погрешность оценки **0.43%**) произнесенных фонем при различных уровнях амплитудного ограничения сигналов и обнаруживает речевой сигнал **даже на фоне мощных шумов**.

Выбор в качестве **материнского комплексного вейвлета Морле** и результаты анализа экспериментальных исследований параметров гласных звуков, полученных с помощью данного вейвлет-преобразования, **позволяют гибко управлять настройкой параметров вейвлета для получения тонкой структуры сигнала конкретного диктора в реальном масштабе времени**.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Величкин, А.И. Передача аналоговых сообщений по цифровым каналам связи. / А.И. Величкин. – М.: Радио и связь, 1983. – 240 с.: ил. (Статистическая теория связи Вып. 19).
2. Бураченок, И. Б. Анализ вейвлет-преобразованием тонкой структуры гласных звуков речевого сигнала / И. Б. Бураченок, В. К. Железняк // Теоретические и прикладные аспекты информационной безопасности : материалы междунар. науч.-практ. конф., Минск, 19 июня 2014 г. / УО «Акад. М-ва внутр. дел Респ. Беларусь»; редкол.: В. Б. Шабанов (отв. ред.) [и др.]. – Минск, 2015. – С. 124-128.
3. Бураченок, И. Б. Обнаружение измерительных сигналов в маскирующих шумах высокого уровня / И. Б. Бураченок, В. К. Железняк, А.Г. Филиппович // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. – 2018. – № 4. – С. 2–9.
4. Бураченок, И. Б. Обнаружение первичных признаков речевого сигнала / И. Б. Бураченок, В. К. Железняк // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. – 2020. – № 12. – С. 2-12.
5. Денисенко, А.Н. Статистическая теория радиотехнических систем. / А.Н. Денисенко. – М.: АРИ, 2007 – 200 с.: ил.
6. Михайлов, В.Г. Измерение параметров речи. / В.Г. Михайлов, Л.В. Златоустова; под ред. М.А. Сапожкова. – М: Радио и связь, 1987 – 168 с.: ил.
7. Стопцов, Н.А. Связь под водой: Справочное пособие. / Н.А.Стопцов, В.И. Бойцов, В.Н. Шелемин. – Л.: Судостроение, 1990. – 248 с.: ил.
8. Фланаган, Дж. Анализ, синтез и восприятие речи / Дж. Фланаган; пер. с англ. под ред. А. А. Пирогова. – М. : Связь, 1968. – 396 с.

**Спасибо за внимание!**

Бураченко Ирина Брониславовна,  
*к.т.н., доцент,  
заведующий кафедрой математики и  
компьютерной безопасности*

*[irina.psu@psu.by](mailto:irina.psu@psu.by)*

Учреждение образования  
«Полоцкий государственный университет  
имени Евфросинии Полоцкой»



# ИСПОЛЬЗОВАНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ В ЗАДАЧАХ ОБРАБОТКИ И РАСПОЗНАВАНИЯ РЕЧИ

Бураченко Ирина Брониславовна,  
*к.т.н., доцент,  
заведующий кафедрой математики и  
компьютерной безопасности*

г. Новополоцк