

# ИСПОЛЬЗОВАНИЕ ДИСКРЕТНОЙ ОЦЕНКИ СПЕКТРА АВТОРЕГРЕССИОННОГО ПРОЦЕССА ДЛЯ ПОЛУЧЕНИЯ ПРИЗНАКОВ В ЗАДАЧАХ АНАЛИЗА ГОЛОСОВОГО СИГНАЛА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

*Лихачёв Д.С., Вашкевич М.И., Петровский Н.А., Азаров И.С.*



*Белорусский государственный университет  
информатики и радиоэлектроники (БГУИР)*

*Кафедра электронных вычислительных средств*



DSPA Conference

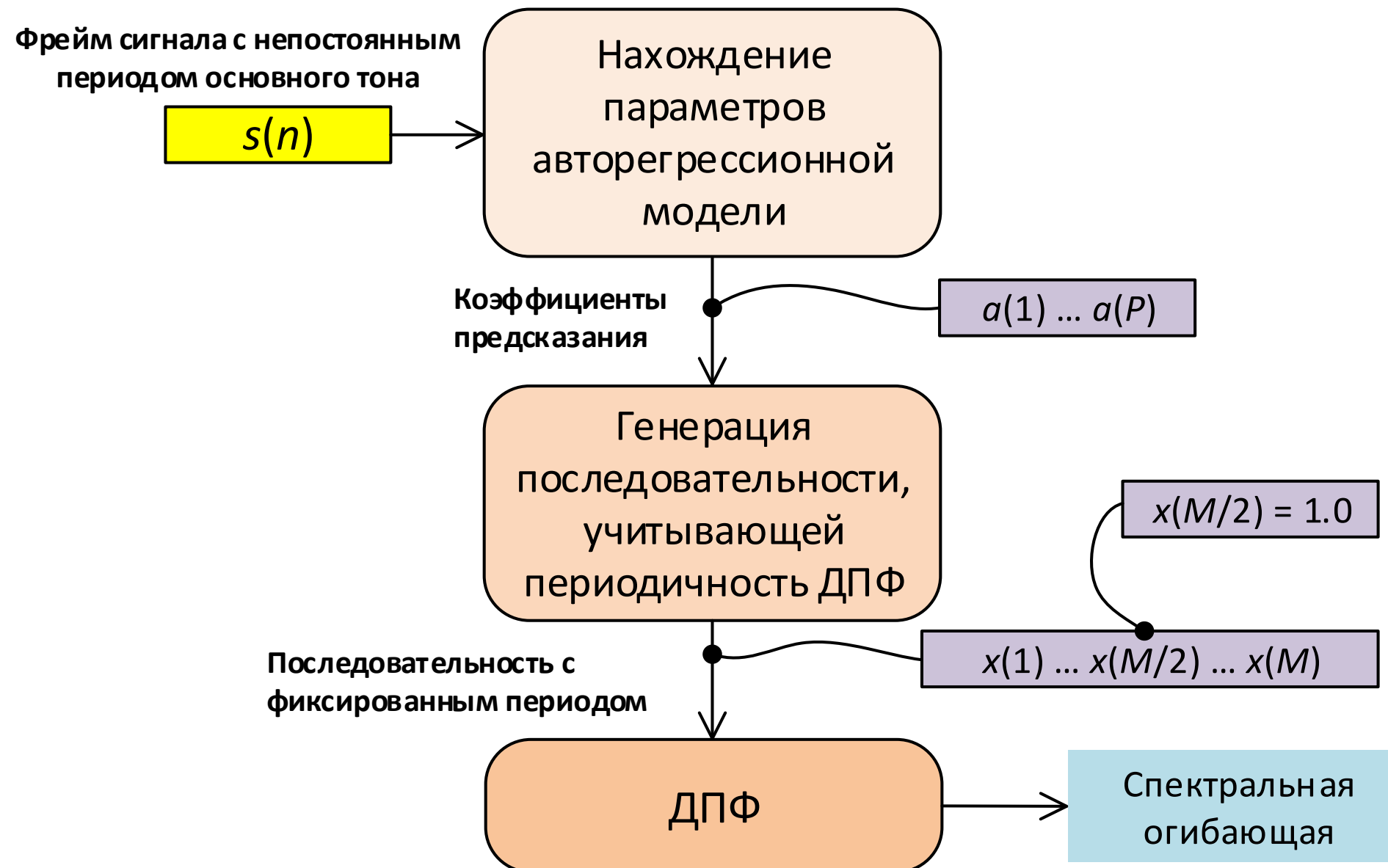
Digital Signal Processing and Its Applications

# Цели и задачи

- ✓ Во многих системах машинного обучения, предназначенных для обработки голосовых и аудио сигналов, для вычисления признаков в частотной области используется ДПФ
- ✓ **Проблема:** спектральная оценка будет точной только при соблюдении условия стационарности анализируемого сигнала и совпадении его периода с размером ДПФ, что на практике не выполняется. Применение временного окна для уменьшения влияния этих факторов приводит к размыванию спектра.
- ✓ В данной работе для получения признаков в задачах анализа и классификации голосового сигнала с помощью машинного обучения предлагается использовать **генеративный метод** спектрального оценивания на основе **дискретной оценки огибающей спектра авторегрессионного процесса** (ОСАП).

# Дискретная оценка ОСАП

**Основная идея** - использовать ДПФ без применения оконной функции для обработки входных данных, обеспечив при этом согласованность периода анализируемого сигнала и размера ДПФ.



# {1} Вычисление дискретной оценки ОСАП

**Шаг 1.** По входному кадру анализируемого сигнала  $x(n)$ ,  $n = \overline{1, N}$  с использованием автокорреляционного метода определяются параметры авторегрессионной модели  $P$ -го порядка:  $a_k$ ,  $k = \overline{1, P}$ .

**Шаг 2.** Линейный предсказатель  $P$ -го порядка для генерируемой дискретной последовательности  $x(m)$  с периодом  $M$ , равным размеру ДПФ:

$$x(m) = - \sum_{k=1}^P a_k x(m - k), \quad m = \overline{1, M + P}, \quad (1)$$

$$\sum_{m=1}^M |x(m)| \neq 0 \quad (2)$$

## {2} Вычисление дискретной оценки ОСАП

Положив  $x(m) = 0$  при  $m = \overline{M+1, M+P}$  и при  $m = \overline{-P+1, 0}$ , можно получить:

$$\begin{pmatrix} -1 & 0 & 0 & \dots & & \\ -a_1 & -1 & 0 & \dots & & \\ -a_2 & a_1 & -1 & \dots & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & & a_P \end{pmatrix} \times \begin{pmatrix} x(1) \\ x(2) \\ x(3) \\ \vdots \\ x(M) \end{pmatrix} = 0 \quad (3)$$

Введя дополнительное ограничение  $x\left(\frac{M}{2}\right) = 1.0$  для выполнения условия (1) и решив полученную из (3) систему линейных уравнений относительно  $x$ , можно найти искомую последовательность  $x(m)$ , которая будет затухать к краям кадра, аналогично оконной функции.

**Шаг 3.** С помощью ДПФ вычисляется дискретный спектр одного периода сгенерированной на шаге 2 последовательности.

## {1} Анализ эффективности представления спектра с помощью ОСАП

Производилось сравнение точности представления спектра *тестового речеподобного сигнала с заданными параметрами* и спектральных огибающих, полученных тремя способами:

- ✓ с использованием дискретной оценки ОСАП
- ✓ ДПФ без применения окна
- ✓ ДПФ с применением окна Ханна

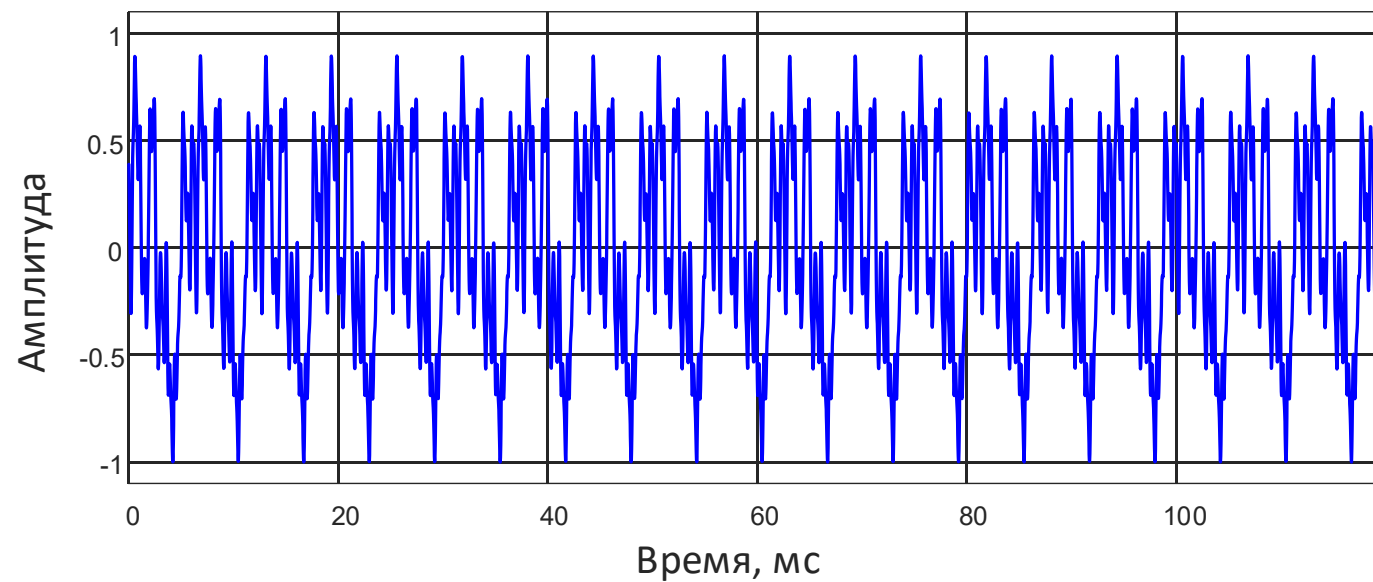
**Критерий точности** представления спектра тестового сигнала – относительная среднеквадратичная ошибка  $SE_{acc}^{i,j}$  между вычисленной спектральной огибающей  $S_{est}^{i,j}(k)$  и спектром тестового сигнала  $S_{test}^i(k)$ :

$$SE_{acc}^{i,j} = \frac{\sum_{k=1}^K \left( S_{test}^i(k) - S_{est}^{i,j}(k) \right)^2}{\sum_{k=1}^K \left( S_{test}^i(k) \right)^2} 100, \%$$

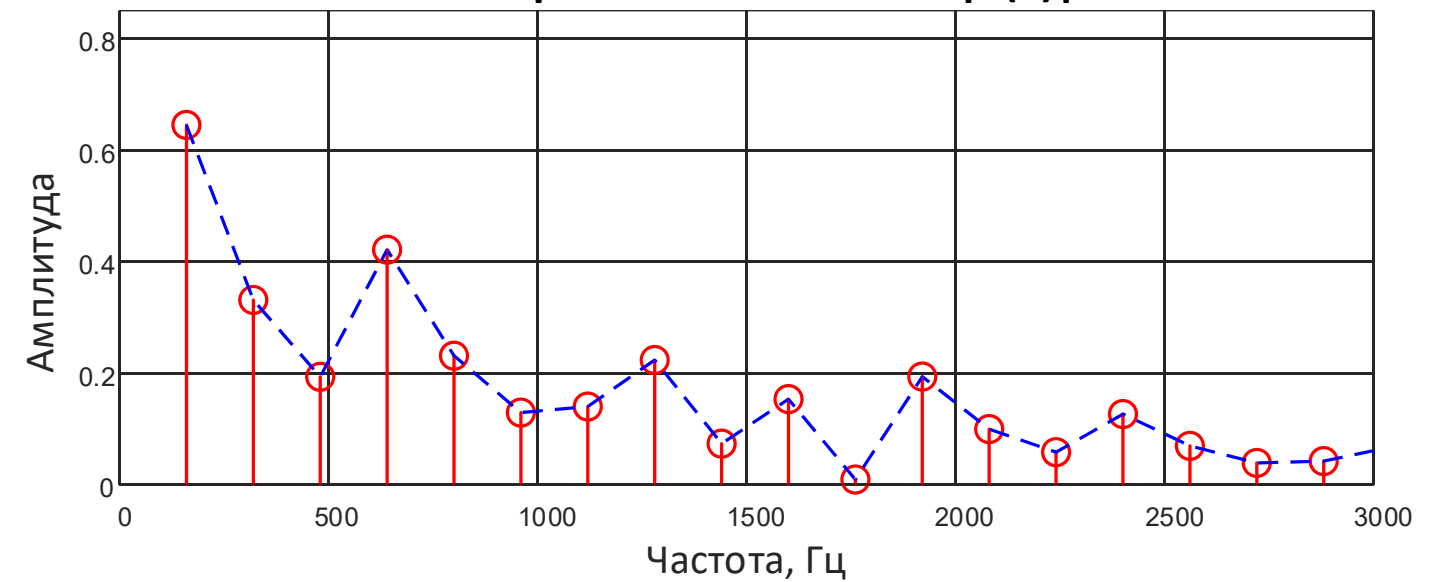
# {1} Анализ эффективности представления спектра с помощью ОСАП

## Тестовый сигнал и соответствующие спектральные огибающие для одного кадра

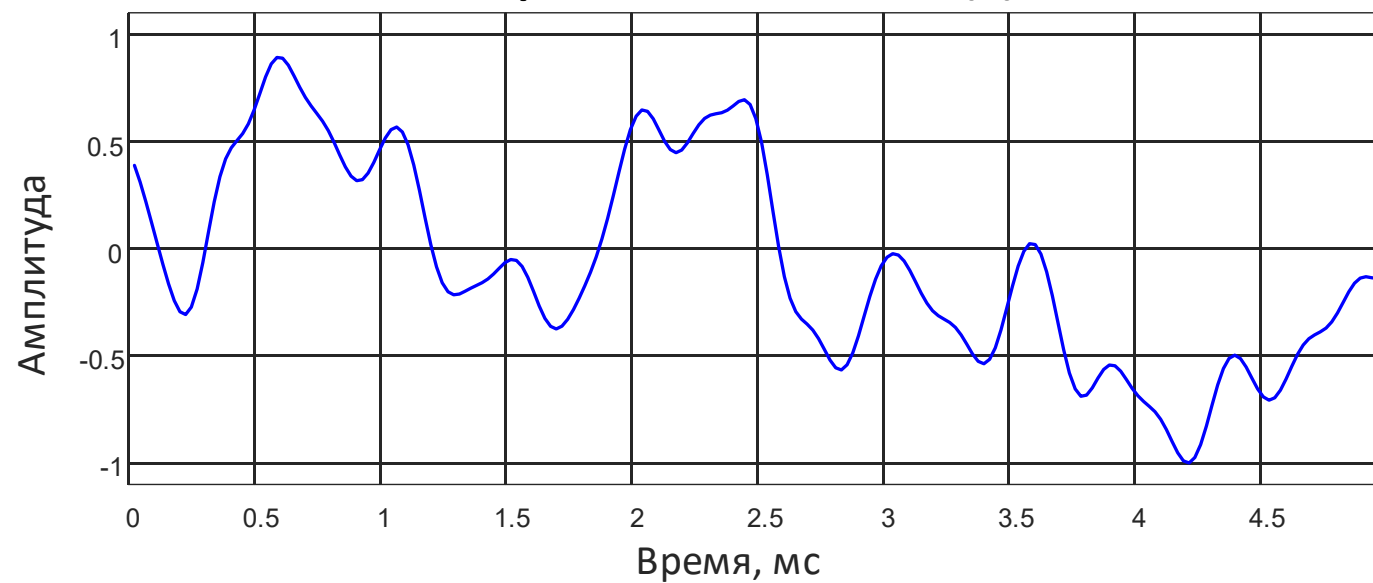
Сегмент тестового сигнала



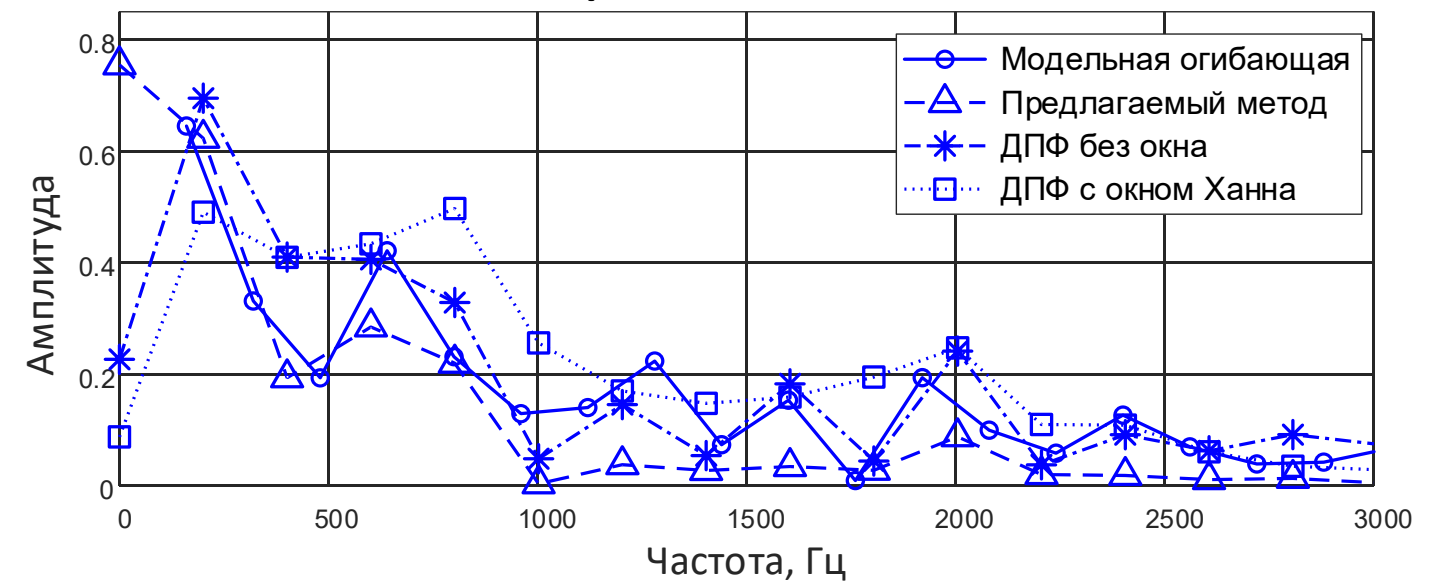
Спектр тестового сигнала  $|S(k)|$



Кадр тестового сигнала  $s(n)$



Спектральные огибающие



## {2} Анализ эффективности представления спектра с помощью ОСАП

Средняя квадратичная ошибка для кадра анализа с разной длительностью, %

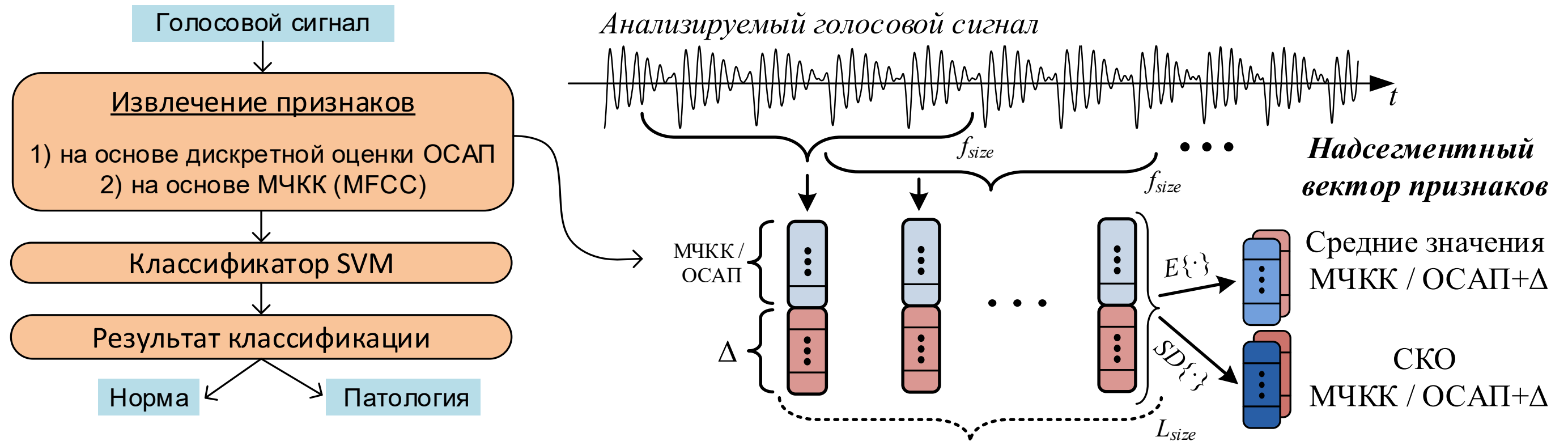
Метод получения спектральной огибающей	Длительность кадра анализа, мс				
	1,5	3,2	5,0	10	20
Предложенный метод	<b>24,8</b>	<b>22,0</b>	<b>14,3</b>	<b>12,2</b>	<b>16,2</b>
ДПФ без применения окна	46,2	33,1	28,1	29,3	42,6
ДПФ с применением окна Ханна	39,4	29,1	19,4	19,8	36,6

*Спектральные огибающие, полученные с помощью метода на основе дискретной оценки ОСАП, более точно по сравнению с остальными методами представляют спектр тестового сигнала с точки зрения минимальной квадратичной ошибки.*



# {1} Исследование эффективности применения признаков на основе ОСАП

## Тестовая система анализа и классификации голосового сигнала



Данная система предназначена для **диагностики БАС по голосу** и выполняет бинарную классификацию входного голосового сигнала по двум категориям: «патология» или «норма»

Для обучения и тестирования системы использовалась база, содержащая **54 образца** голосовых сигналов с протяжным звуком /a/.

## {2} Исследование эффективности применения признаков на основе ОСАП

**Метод перекрестной проверки LOSO** (Leave-One-Subject-Out):

- итеративное тестирование
- количество итераций равнялось количеству дикторов
- на каждой итерации в тестовый набор включались данные, соответствующие одному диктору, а в обучающий – все оставшиеся
- на каждой итерации определялась сбалансированная точность классификации  $BA$ :

$$BA = 0,5 \left( \frac{TP}{TP + FN} \right) + \left( \frac{TN}{TN + FP} \right),$$

где  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  – количество истинно положительных, истинно отрицательных, ложноположительных и ложноотрицательных результатов классификации

*В качестве оценки эффективности классификации использовалось среднее значения сбалансированной точности по всем итерациям  $\overline{BA}$*

### {3} Исследование эффективности применения признаков на основе ОСАП

Максимальная средняя сбалансированная точность классификации при использовании МЧКК признаков в зависимости от количества частотных отсчетов  $M$  и длительности кадра анализа

Длительность кадра анализа, мс	Сбалансированная точность классификации $\overline{BA}$ , %										
	$M \leq 10$	$M \leq 15$	$M \leq 20$	$M \leq 25$	$M \leq 30$	$M \leq 35$	$M \leq 40$	$M \leq 45$	$M \leq 50$	$M \leq 55$	$M \leq 60$
10	69,3	70,5	71,7	71,7	<b>73,2</b>	67,2	70,3	72,4	71,9	70,7	69,7
15	70,1	70,2	70,1	70,1	69,1	71,6	71,8	73,0	<b>73,7</b>	69,9	70,8
20	71,3	68,3	71,3	72,5	73,9	72,6	70,8	<b>74,4</b>	71,2	72,1	72,1
25	72,5	68,8	72,2	72,7	<b>75,7</b>	69,3	70,4	72,3	69,4	72,0	70,8
30	73,0	67,5	74,3	<b>74,7</b>	70,9	69,7	70,2	70,1	70,3	72,2	69,7
35	74,3	67,3	69,3	75,4	71,2	69,7	75,0	69,6	72,3	<b>77,3</b>	69,5
40	73,9	67,1	70,9	<b>75,0</b>	67,9	69,3	71,4	68,8	72,5	71,3	72,8

## {4} Исследование эффективности применения признаков на основе ОСАП

Максимальная средняя сбалансированная точность классификации при использовании предлагаемых признаков в зависимости от количества частотных отсчетов  $M$  и длительности кадра анализа

Длительность кадра анализа, мс	Сбалансированная точность классификации $\overline{BA}$ , %										
	$M \leq 10$	$M \leq 15$	$M \leq 20$	$M \leq 25$	$M \leq 30$	$M \leq 35$	$M \leq 40$	$M \leq 45$	$M \leq 50$	$M \leq 55$	$M \leq 60$
10	72,5	81,0	<b>81,6</b>	80,4	74,2	74,6	73,1	78,5	78,2	78,5	75,7
15	69,2	79,0	<b>80,3</b>	79,5	76,6	74,4	74,7	76,6	78,5	77,6	78,9
20	67,4	78,6	<b>80,5</b>	79,4	77,9	75,2	74,8	77,9	79,7	79,4	77,7
25	67,1	77,0	<b>81,5</b>	77,5	77,7	75,4	73,3	75,6	78,1	78,6	77,2
30	67,0	74,4	<b>82,7</b>	76,2	77,4	73,9	72,7	74,1	78,6	77,5	77,1
35	66,8	74,6	<b>81,0</b>	76,2	78,7	74,1	72,2	73,7	77,5	77,2	76,8
40	66,2	74,5	<b>81,5</b>	75,7	77,5	73,6	71,7	72,9	76,5	78,3	75,6

# Выводы

- ✓ В работе предложен подход для получения спектральных признаков на основе дискретной оценки ОСАП, которые является альтернативой классическим мел-спектральным признакам
- ✓ Результаты проведенных экспериментов для определения патологических изменений голоса свидетельствуют, что представленный метод позволяет более эффективно извлекать полезную для классификации информацию из имеющихся обучающих данных и ощутимо повышать точность классификации.
- ✓ Использование предлагаемых признаков в системе диагностики БАС позволило повысить сбалансированную точность классификации на величину до 8%.
- ✓ Предложенный подход имеет хорошую перспективу применения для задач обработки речевых и голосовых сигналов, особенно в системах машинного обучения с малым объемом обучающих данных

**Спасибо за внимание!**