

# МЕТОД ПОНИЖЕНИЯ РАЗМЕРНОСТИ ПРОСТРАНСТВА ПРИЗНАКОВ НА ОСНОВЕ LASSO-РЕГРЕССИИ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО РЕЧИ

*маг. Краснопрошин Д.В., проф. Вашкевич М.И.*



*Белорусский государственный университет  
информатики и радиоэлектроники (БГУИР)*

*Кафедра электронных вычислительных средств*



# План доклада

1. Цель работы
2. Процесс разработки системы распознавания эмоций
3. Процедуры отбора признаков
4. Метод LASSO
5. Набор данных и описание эксперимента
6. Результаты и заключение

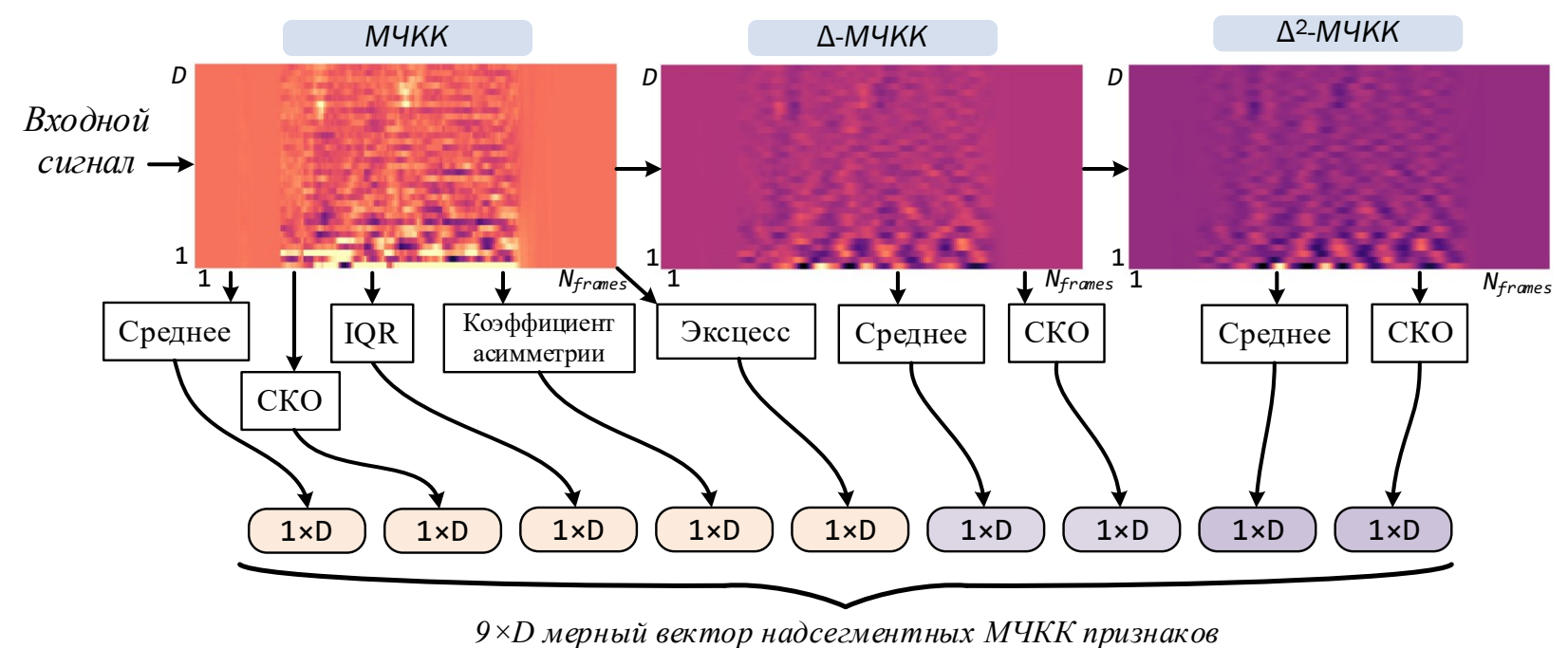
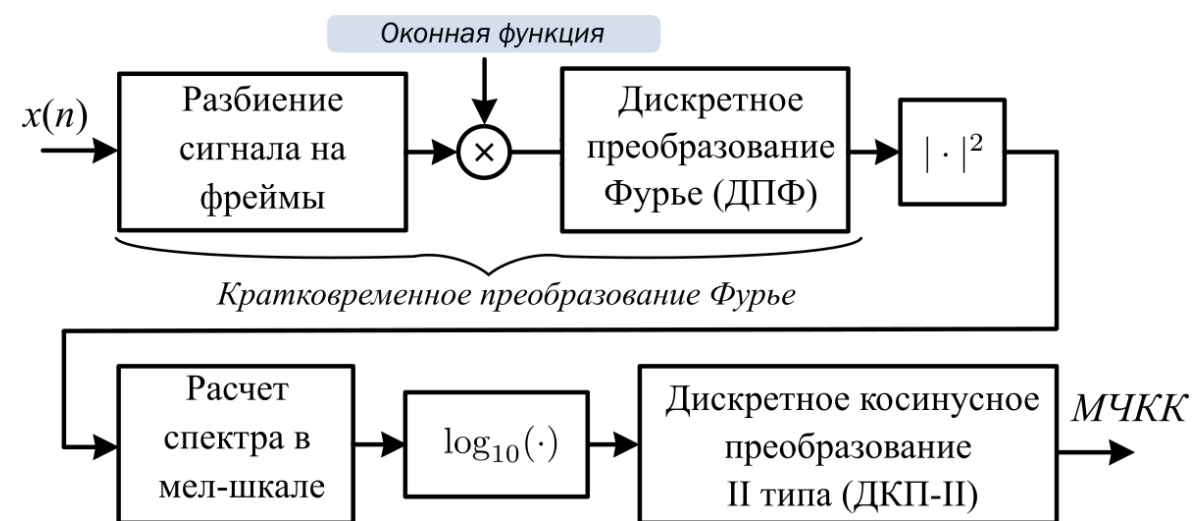
## Цель доклада

*Представить метод понижения размерности пространства признаков, основанный на применении LASSO-регрессии, для повышения эффективности распознавания эмоций по речи.*

# Разработка системы распознавания эмоций

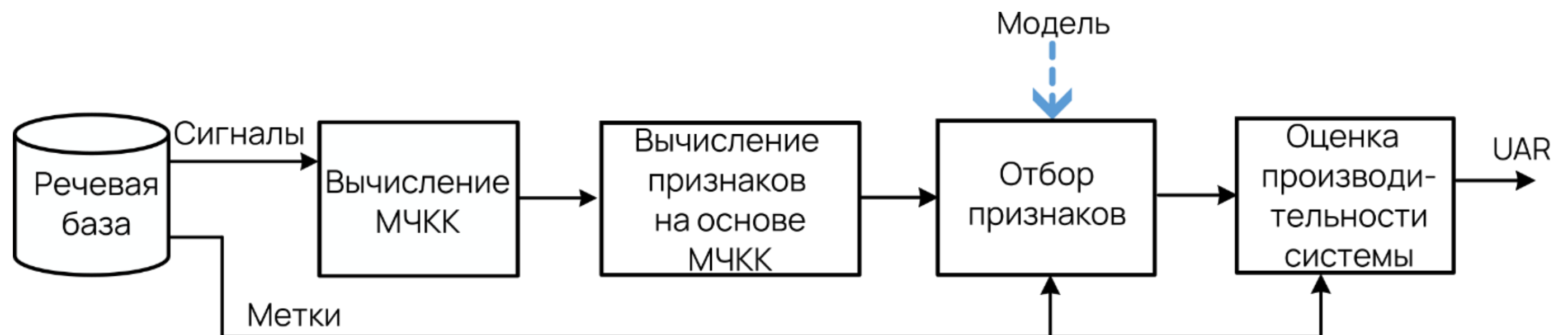
## Речевы признаки

Мел-частотные кепстральные коэффициенты (МЧКК),  $\Delta$  МЧКК,  $\Delta^2$  МЧКК, межквартильный размах, коэффициент асимметрии и эксцесс.



# Процесс разработки системы распознавания эмоций

- В процессе разработки системы выполняется **отбор признаков**, которые способствуют повышению производительности системы распознавания эмоций. Данный этап выполнялся с учетом модели классификатора, который будет использоваться в дальнейшем в системе распознавания эмоций.
- Качество модели оценивалось с помощью UAR (Unweighted Average Recall) и перекрестной проверки.



# Процедура отбора признаков

Разработан **алгоритм отбора признаков** для задачи многоклассовой классификации, основанный на использовании LASSO-регрессии:

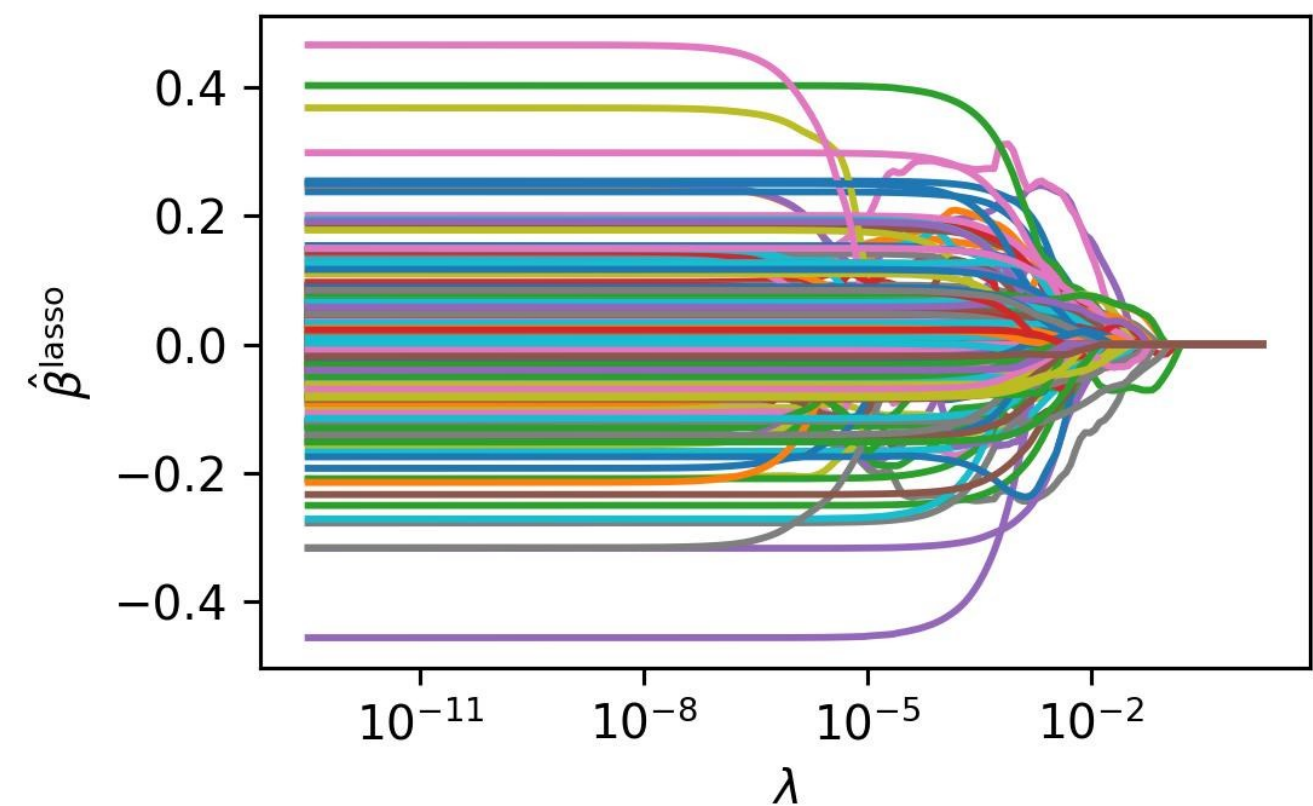
- 1) Задача классификации на  $K$ -классов  $\{C_1, C_2, \dots, C_K\}$  заменяется на  $K$  задач бинарной классификации, по схеме «один против всех».
- 2) Для решения задач бинарной классификации применяется метод отбора признаков на основе LASSO. В результате получается  $K$  подмножеств признаков из их исходного полного набора.
- 3) Для каждого из  $K$  наборов признаков, полученных на шаге 2, применяется метод пошагового исключения переменных (англ. *BSS – backward-stepwise selection*). Этот этап позволяет выбрать оптимальное подмножество наиболее значимых признаков, снижая размерность данных и устраняя избыточность.
- 4) Полученные на шаге 3  $K$  наборов признаков объединяются для формирования итогового набора признаков.

# Применение LASSO для отбора признаков в задаче бинарной классификации

Метод LASSO основан на решении задачи линейной регрессии:

$$\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (1)$$

$N$  – число примеров в обучающем наборе,  $y_i$  – метка класса  $i$ -го образца,  $x_{ij}$  –  $j$ -й признак  $i$ -го образца,  $\beta_j$  –  $j$ -й коэффициент регрессии,  $\lambda$  – параметр регуляризации. Решение (1) при достаточно большом  $\lambda$  приводит к тому, что часть коэффициентов  $\beta$  становятся в точности нулевыми.

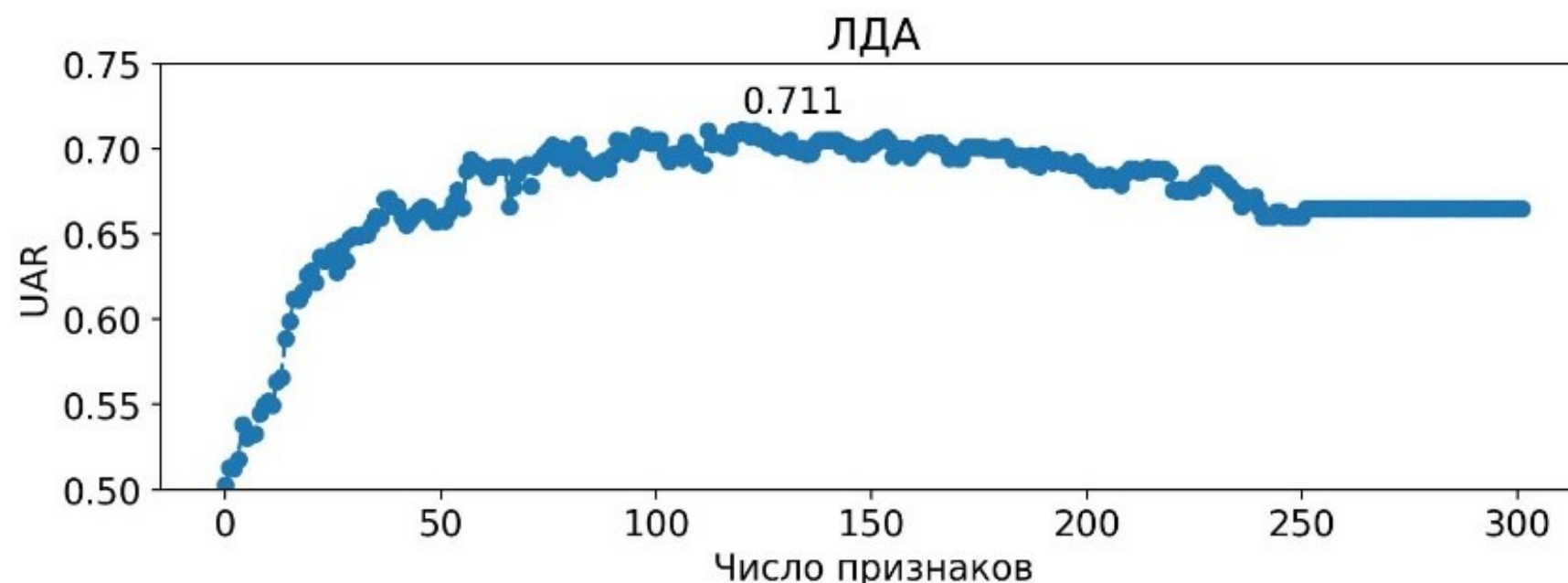




# Применение LASSO для отбора признаков в задаче бинарной классификации

## Отбор подмножества признаков по результатам ранжирования с на основе LASSO-регрессии

После ранжирования признаков выполняется оценка точности классификации (UAR) при использовании возрастающего набора признаков. В ходе этой процедуры определяется число признаков, которое обеспечивает наибольшую точность классификации.





# Набор данных и извлечение признаков

## Речевая база

- RAVDES (Ryerson Audio-Visual Database of Emotional Speech and Song), 8 эмоциональных состояний (нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение), 12 мужчин и 12 женщин, 1440 аудиофайлов в формате wav (16 бит, 48 кГц)
- RAVDESS содержит эмоциональные состояния: нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение.

Речевые признаки рассчитывались на основании мел-частотных кепстральных коэффициентов (МЧКК). В итоговый набор *исходных признаков* были включены среднее значение МЧКК (34 признака), среднеквадратичное отклонение МЧКК (34 признака), среднее от первой и второй производных от МЧКК (68 признаков), их среднеквадратическое отклонение (68), а также коэффициент асимметрии, эксцесс и межквантильный размах (по 34 признака для каждой характеристики соответственно). Таким образом, для каждого аудиофайла мы получаем 306-компонентный вектор надсегментных признаков МЧКК.

# Тестирование классификатора

Для тестирования классификатора использовался метод перекрестной проверки по  $k$ -блокам (*k-fold cross-validation*)

В данной работе использовалась схема разбиения на блоки из работы [1] (в скобках указаны номера актеров):

- блок 0: (2, 5, 14, 15, 16);
- блок 1: (3, 6, 7, 13, 18);
- блок 2: (10, 11, 12, 19, 20);
- блок 3: (8, 17, 21, 23, 24);
- блок 4: (1, 4, 9, 22).

[1] *Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning* / C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J.M. Montero, F. Fernández-Martínez // *Sensors*. – 2021. – vol. 21. – pp. 1 – 29.

# Результаты экспериментальных исследований

Исследовалось применение классификаторов на основе

- линейного дискриминантного анализа (ЛДА)
- машины опорных векторов (МОВ).

Оценка производительности выполнялась с использованием невзвешенной средней полноты – UAR (*Unweighted Average Recall*).

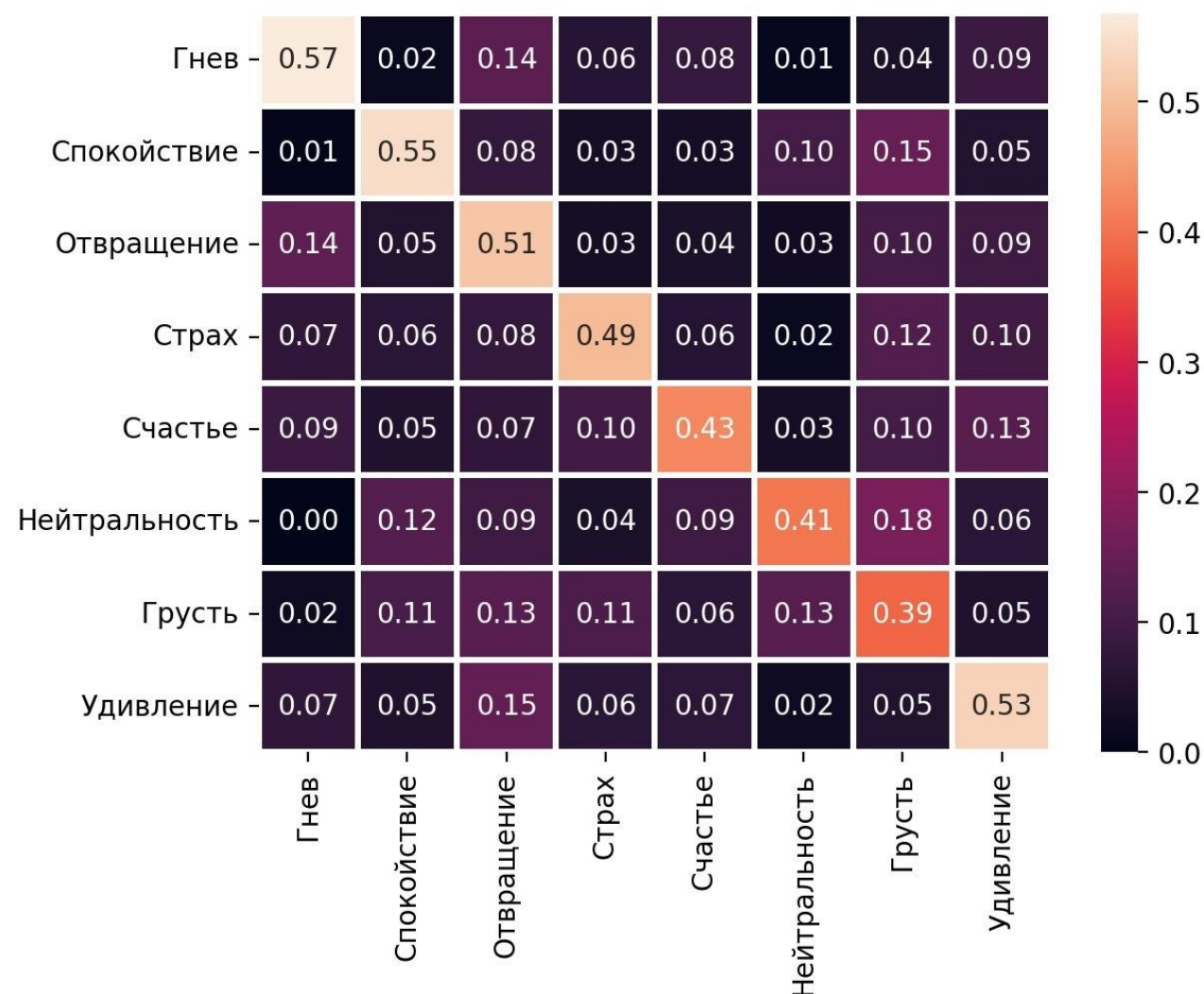
# Результаты экспериментальных исследований

Классификатор	Полный набора признаков (306)	Отобранный набор признаков
ЛДА	0,460	<b>0,484</b> (205 признаков)
МОВ	0.456	<b>0.473</b> (208 признаков)

- После процедуры отбора признаков из изначального набора, состоящего из 306 характеристик, оставлено 205 признаков при повышении точности классификатора в случае с ЛДА и 208 признаков в случае с МОВ.
- Результат демонстрирует значительное уменьшение размерности признакового пространства.
- Одновременно удалось сохранить высокий уровень точности.

# Результаты экспериментальных исследований

## Матрица спутанности



**UAR = 0,484** (205 признаков)

## Сравнение с [1]

Авторы применили Transfer Learning используя AlexNet и CNN-14 для извлечения признаков. MOB натренированные на признаках извлеченных AlexNet и CNN-14 продемонстрировали **UAR  $43.32 \pm 2.56$**  и  **$39.73 \pm 2.53$**  соответственно.

<sup>1</sup> [1] *Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning* / C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J.M. Montero, F. Fernández-Martínez // *Sensors*. – 2021. – vol. 21. – pp. 1 – 29.

# Заключение

*В работе предложен метод понижения размерности пространства признаков, основанный на применении LASSO-регрессии, для повышения эффективности распознавания эмоций по речи.*

В работе выполнено сравнение качества классификации модели с использованием исходного набора признаков, и набора признаков полученных с помощью разработанного метода понижения размерности признакового пространства. Благодаря новому методу **удалось снизить размерность примерно на 30% и одновременно удалось сохранить высокий уровень классификационной точности, что подчеркивает эффективность примененной методологии отбора признаков.**