



<http://dx.doi.org/10.35596/1729-7648-2024-22-3-93-100>

Оригинальная статья  
Original paper

УДК 004.934.2+534.784

## МЕТОД РАСПОЗНАВАНИЯ ЭМОЦИЙ В РЕЧЕВОМ СИГНАЛЕ С ИСПОЛЬЗОВАНИЕМ МАШИНЫ ОПОРНЫХ ВЕКТОРОВ И НАДСЕГМЕНТНЫХ АКУСТИЧЕСКИХ ПРИЗНАКОВ

Д. В. КРАСНОПРОШИН, М. И. ВАШКЕВИЧ

*Белорусский государственный университет информатики и радиоэлектроники  
(г. Минск, Республика Беларусь)*

*Поступила в редакцию 19.12.2023*

© Белорусский государственный университет информатики и радиоэлектроники, 2024  
Belarusian State University of Informatics and Radioelectronics, 2024

**Аннотация.** Исследована задача распознавания эмоций в речевом сигнале с использованием мел-частотных кепстральных коэффициентов при помощи классификатора на основе метода опорных векторов. При проведении экспериментов применялся набор данных RAVDESS. Предложена модель, которая использует 306-компонентный вектор надсегментных признаков в качестве входных данных для классификатора на основе метода опорных векторов. Оценка качества модели проводилась с помощью невзвешенного среднего значения полноты (UAR). Рассмотрено применение в классификаторе на основе метода опорных векторов в качестве ядра линейной, полиномиальной и радиальной базисной функций. Исследовано использование разных размеров фрейма анализа сигнала (от 23 до 341 мс) на этапе извлечения мел-частотных кепстральных коэффициентов. Результаты исследований выявили значительную точность полученной модели (UAR = 48 %). Предлагаемый подход демонстрирует потенциал для таких приложений, как голосовые помощники, виртуальные агенты и диагностика психического здоровья.

**Ключевые слова:** голосовой сигнал, мел-частотные кепстральные коэффициенты, извлечение аудиопризнаков, распознавание, машинное обучение.

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Для цитирования.** Краснопрошин, Д. В. Метод распознавания эмоций в речевом сигнале с использованием машины опорных векторов и надсегментных акустических признаков / Д. В. Краснопрошин, М. И. Вашкевич // Доклады БГУИР. 2024. Т. 22, № 3. С. 93–100. <http://dx.doi.org/10.35596/1729-7648-2024-22-3-93-100>.

## SPEECH EMOTION RECOGNITION METHOD BASED ON SUPPORT VECTOR MACHINE AND SUPRASEGMENTAL ACOUSTIC FEATURES

DANIIL V. KRASNOPROSHIN, MAXIM I. VASHKEVICH

*Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)*

*Submitted 19.12.2023*

**Abstract.** The problem of recognizing emotions in a speech signal using mel-frequency cepstral coefficients using a classifier based on the support vector machine has been studied. The RAVDESS data set was used in the experiments. A model is proposed that uses a 306-component suprasegmental feature vector as input to a support vector machine classifier. Model quality was assessed using unweighted average recall (UAR). The use of linear, polynomial and radial basis functions as a kernel in a classifier based on the support vector machine is considered. The use of different signal analysis frame sizes (from 23 to 341 ms) at the stage of extracting mel-frequency cepstral coefficients was investigated. The research results revealed significant accuracy of the resulting mo-

del (UAR = 48 %). The proposed approach shows potential for applications such as voice assistants, virtual agents, and mental health diagnostics.

**Keywords:** voice signal, mel-frequency cepstral coefficients, audio feature extraction, recognition, machine learning.

**Conflict of interests.** The authors declare no conflict of interests.

**For citation.** Krasnoproshin D. V., Vashkevich M. I. (2024) Speech Emotion Recognition Method Based on Support Vector Machine and Suprasegmental Acoustic Features. *Doklady BGUIR*. 22 (3), 93–100. <http://dx.doi.org/10.35596/1729-7648-2024-22-3-93-100> (in Russian).

## Введение

Область распознавания эмоций по речи быстро развивается благодаря росту производительности вычислительных систем и широкому интересу к этому вопросу исследователей в сфере психологии, психиатрии и информатики [1, 2]. Эмоции часто влияют на процессы принятия решений, поэтому их распознавание может представлять интерес для построения более эффективного общения, включая диалоговые системы (голосовые помощники, чат-боты). Задача распознавания негативных эмоций, таких как стресс, гнев, усталость, является важным аспектом с точки зрения обеспечения безопасности дорожного движения при использовании интеллектуальных транспортных средств, поскольку позволяет им реагировать на эмоциональное состояние водителя [3]. В сфере взаимодействия человека и компьютера точное распознавание эмоций по речи является ключевым фактором. В статье рассматривается задача определения эмоций на основе анализа речевого сигнала. Одна из основных проблем данного подхода связана с определением набора признаков, эффективно описывающих эмоциональное состояние человека [1, 4–6].

При построении системы распознавания эмоций предлагается использовать мел-частотные кепстральные коэффициенты (МЧКК) [7] для получения признаков и метод опорных векторов (МОВ) [7] в качестве классификатора. Признаки на основе МЧКК широко применяются в системах распознавания эмоций по речи [1, 5]. МЧКК воспроизводят реакцию слуховой системы человека на звук, улавливая соответствующую акустическую информацию [8]. Формируя представление аудиосигнала в частотной области, МЧКК уменьшают размерность данных, сохраняя при этом важные особенности, что делает их пригодными для алгоритмов машинного обучения, таких как МОВ. В свою очередь, МОВ является простым и надежным подходом к задаче классификации, который обеспечивает адаптируемость к многомерным пространствам признаков. МОВ основан на принципе поиска оптимальной гиперплоскости, максимально разделяющей разные классы в пространстве признаков [9]. В контексте распознавания эмоций по речи это означает, что МОВ может эффективно дифференцировать эмоциональные состояния [5], а также учитывать нелинейные отношения с помощью функций ядра (kernel function), что позволяет улавливать сложные закономерности в речевых данных.

## Извлечение речевых признаков

На начальном этапе в системе распознавания эмоций по речи выполняется предварительная обработка входных аудиоданных [1, 5]. Речевые признаки рассчитываются на основании МЧКК [8]. Расчет МЧКК относится к методам кратковременного анализа речевого сигнала, которые предполагают разбиение сигнала на фреймы (короткие сегменты). Считается, что в интервале от 10 до 30 мс голосовой сигнал можно считать стационарным. На рис. 1 представлена схема вычисления МЧКК.

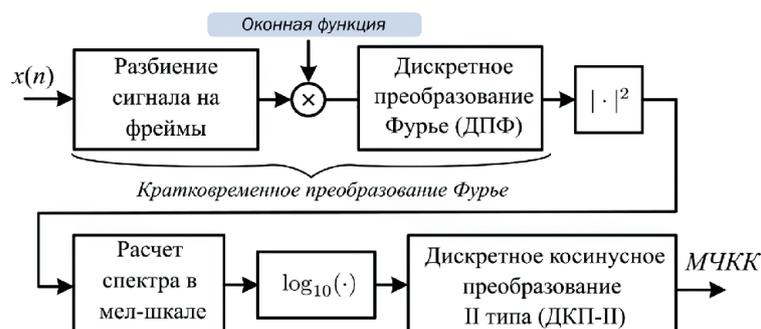


Рис. 1. Схема вычисления мел-частотных кепстральных коэффициентов  
Fig. 1. Scheme for calculating mel-frequency cepstral coefficients

Согласно рис. 1, процесс извлечения МЧКК включает следующие шаги:

- 1) вычисление кратковременного преобразования Фурье (КВПФ) и нахождение квадрата модуля КВПФ для получения спектрограммы сигнала;
- 2) вычисление мел-спектрограммы (энергия сигнала из шкалы герц переводится в мел-шкалу, отражающую свойства человеческого слуха);
- 3) взятие логарифма от энергии сигнала в мел-частотных полосах;
- 4) применение декоррелирующего преобразования, в качестве которого используется дискретное косинусное преобразование II типа (ДКП-II).

В качестве иллюстрации на рис. 2 показан пример речевого сигнала, выражающего эмоцию гнева.

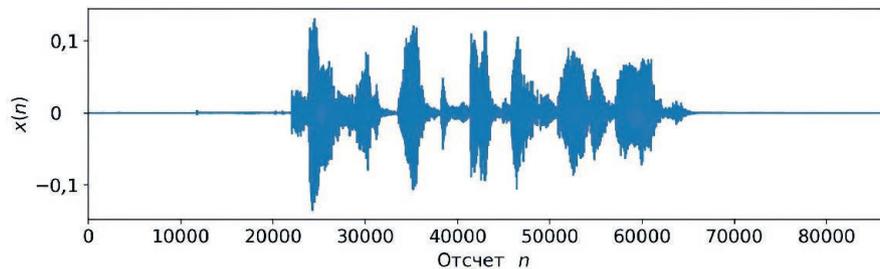


Рис. 2. Представление речевого сигнала, выражающего гнев  
Fig. 2. Representation of the speech signal expressing anger

На рис. 3 показан результат вычисления КВПФ и мел-спектрограммы сигнала, представленного на рис. 1. На рис. 4 изображена временная последовательность МЧКК, рассчитанная для сигнала на рис. 1.

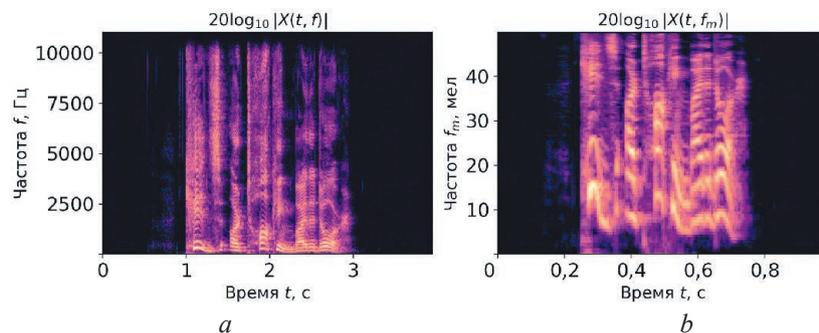


Рис. 3. Речевой сигнал, выражающий гнев: *a* – спектрограмма кратковременного преобразования Фурье; *b* – мел-спектрограмма

Fig. 3. Speech signal expressing anger: *a* – short time Fourier transform spectrogram; *b* – mel-spectrogram

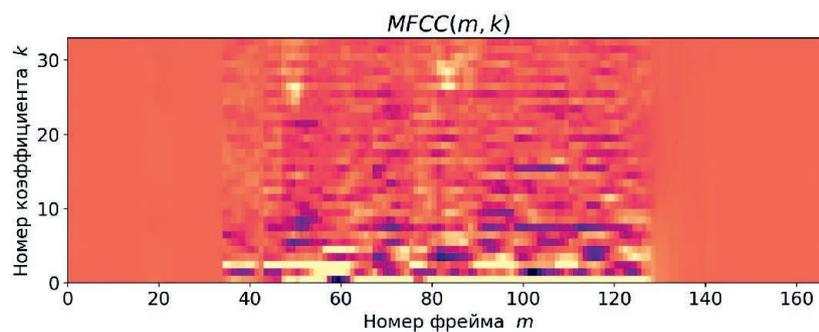


Рис. 4. Временная последовательность мел-частотных кепстральных коэффициентов  
Fig. 4. Time-sequence of mel-frequency cepstral coefficients

В экспериментах использовались речевые сигналы с частотой дискретизации 48 кГц. Для обработки аудиосигналов применялась библиотека librosa, написанная на языке Python. КВПФ рассчитывалось для различных размеров фреймов  $N = \{1024, 2048, 4096, 8192\}$ . Перекрывание между фреймами принималось равным 50 % от размера фрейма. Из каждого фрейма извлекался век-

тор МЧКК размером  $D$ . После обработки одного аудиофайла получается матрица  $MFCC(k, m)$  размером  $D \times N_{frames}$ , где  $N_{frames}$  – количество фреймов. Таким образом,  $m$ -й столбец матрицы представляет собой вектор МЧКК, вычисленный для временного фрейма с номером  $m$ .

Для получения единого вектора признаков для каждого аудиофайла вычисляются средние  $meanMFCC(k)$  ( $D$  признаков) и среднеквадратичные (СКО) отклонения,  $SDMFCC(k)$  ( $D$  признаков) для МЧКК в матрице  $MFCC$  вдоль оси времени. СКО может быть полезно для оценки степени изменчивости в МЧКК во времени. По матрице МЧКК также производилось извлечение первой и второй производных (первой и второй разностей).

Первая производная МЧКК ( $\Delta$ -МЧКК) представляет собой скорость изменения каждого коэффициента во времени. Таким образом,  $\Delta$ -МЧКК отражает быстроту изменения спектральных характеристик звука во времени, помогает в выявлении моментов, когда звук становится более высокочастотным или громким, что может быть полезным для распознавания звуковых событий. Вторая производная МЧКК ( $\Delta^2$ -МЧКК) представляет собой ускорение изменения каждого коэффициента во времени. Принятие во внимание  $\Delta^2$ -МЧКК позволяет более полно описать характер спектральных изменений речевого сигнала во времени.

Известно, что применение производных МЧКК дает возможность улучшить производительность систем распознавания, классификации и детектирования речевых сигналов [1, 5, 10]. Они могут использоваться для выделения важных характеристик аудиосигнала, таких как изменения в тональности, интонации и ритме речи. Это делает их полезными в приложениях по распознаванию речи, детектированию звуковых событий и музыкальному анализу.

Для полученных матриц  $\Delta$ -МЧКК и  $\Delta^2$ -МЧКК рассчитывали их средние значения и СКО, что в итоге дало  $4D = 136$  дополнительных признаков. На основании матрицы МЧКК также были рассчитаны коэффициент асимметрии (skewness), эксцесс (kurtosis) и межквартильный размах (Interquartile Range, IQR). Коэффициент асимметрии – это мера степени асимметрии распределения случайной величины. Она показывает, насколько сильно и в какую сторону смещено распределение относительно своего среднего значения. В контексте анализа МЧКК коэффициент асимметрии может дать информацию о форме распределения и подчеркнуть определенные аспекты вариабельности звукового сигнала. Коэффициент асимметрии вычислялся для каждого МЧКК

$$SkMFCC(k) = \left( \frac{1}{N_{frames}} \right) \sum_{m=0}^{N_{frames}} \frac{(MFCC(k, m) - meanMFCC(k))^3}{SDMFCC(k)^3}. \quad (1)$$

Эксцесс – мера формы распределения случайной величины, которая показывает, насколько оно остроконечное или плоское по сравнению с нормальным распределением. Для вычисления эксцесса использовали формулу моментов

$$kurtosisMFCC(k) = \left( \frac{1}{N_{frames}} \right) \sum_{m=0}^{N_{frames}} \frac{(MFCC(k, m) - meanMFCC(k))^4}{SDMFCC(k)^4 - 4}. \quad (2)$$

Межквартильный размах – это мера разброса данных, которая используется для измерения разницы между верхним и нижним квартилями. Она показывает разброс значений в данных. При вычислении IQR выполняются шаги:

- 1) упорядочиваются данные по возрастанию;
- 2) находится значение первого квартиля ( $Q_1$ ), которое отделяет нижние 25 % наблюдений;
- 3) находится значение третьего квартиля ( $Q_3$ ), которое отделяет нижние 75 %;
- 4) вычисляется IQR как разница между значениями  $Q_3$  и  $Q_1$

$$IQR = Q_3 - Q_1. \quad (3)$$

В контексте решаемой задачи IQR может быть полезен для измерения разброса в параметрах МЧКК. Результирующая схема формирования вектора признаков приведена на рис. 5. В итоговый набор признаков были включены среднее значение МЧКК ( $D$  признаков), среднеквадратичное отклонение МЧКК ( $D$  признаков), среднее от первой и второй производных от МЧКК ( $2D$  признаков), их среднеквадратическое отклонение ( $2D$ ), а также коэффициент асимметрии,

эксцесс и межквантильный размах (по  $D$  признаков для каждой характеристики соответственно). Таким образом, для каждого аудиофайла получается  $9D$ -компонентный вектор надсегментных признаков МЧКК.

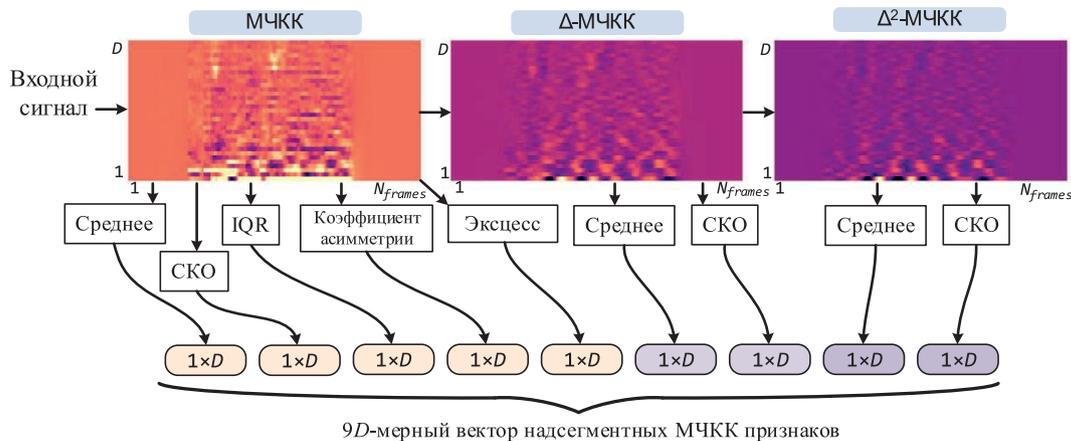


Рис. 5. Схема формирования вектора признаков  
Fig. 5. The scheme of feature vector extraction

### Речевая база и описание эксперимента

При проведении исследования применяли набор данных Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [11]. RAVDESS содержит записи 24 актеров (12 мужчин, 12 женщин). Каждый актер произнес 104 различных сообщения (60 речевых высказываний и 44 песенных). В экспериментах использовалась только часть базы RAVDESS, содержащая речевые высказывания – 1440 файлов в формате wav (16 бит, 48 кГц): 60 записей на каждого из 24-х актеров. Речевые эмоции включали нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение. Все эмоциональные состояния, кроме нейтрального, озвучивались на двух уровнях эмоциональной громкости (нормальном и повышенном). Актеры повторяли каждую вокализацию дважды.

Для классификации в системе распознавания эмоционального состояния по речи применяется МОВ. Данный метод выполняет классификацию путем построения разделяющей гиперплоскости, которая оптимально разделяет данные на классы. В МОВ предусмотрено использование функции ядра для преобразования исходного набора данных в многомерное пространство признаков. И уже в новом пространстве признаков выполняется построение оптимальной разделяющей гиперплоскости. В исследовании использовались известные ядерные функции: линейная, полиномиальная и гауссово ядро с радиальной базисной функцией (РБФ) [9].

Построение классификатора на опорных векторах с применением перечисленных выше ядер осуществлялось с помощью библиотеки sklearn языка Python. Для тестирования классификатора использовался метод перекрестной проверки по  $k$  блокам ( $k$ -fold cross-validation) [9], который заключается в следующем.

1. Набор разбивается на  $k$  блоков.
2. В цикле для  $i = 1, 2, \dots, k$  выполняются следующие операции:
  - блок  $i$  устанавливается в качестве тестового набора данных (test data);
  - оставшиеся блоки принимаются как тренировочные данные (train data);
  - выполняется обучение модели классификатора на тренировочных и оценивается ее эффективность на тестовых данных;
  - сохранение результатов классификации для данных из тестового набора;
  - сброс параметров модели до исходного состояния для следующей итерации.
3. Расчет оценки эффективности модели на основе сохраненных результатов классификации тестовых данных.

Полученные данные, согласно схеме, предложенной в [12], разбивали на блоки следующим образом (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).

Выбранная стратегия заключается в том, что каждый блок должен содержать одинаковое количество случайно выбранных образцов для каждого класса. При этом должно выполняться условие, что каждый актер представлен либо в обучающей, либо в валидационной выборке, но не в обеих. Для оценки качества модели вычисляли среднее арифметическое (невзвешенное) полноты (unweighted average recall, UAR). UAR – это показатель, используемый для измерения общей производительности модели многоклассовой классификации, вычисляет средний уровень запоминания по всем классам, придавая каждому классу одинаковую важность без учета классового дисбаланса:

$$UAR = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{A_{ii}}{\sum_{j=1}^{N_c} A_{ij}}, \quad (4)$$

где  $A$  – матрица спутанности (confusion matrix);  $N_c$  – количество классов.

Значение UAR находится в диапазоне от 0 до 1. Эксперимент проводился в три этапа:

- 1) подготовка обучающей выборки;
- 2) обучение и тестирование классификатора с использованием различных функций ядра и разных размеров длины фрейма анализа речевых сигналов;
- 3) оценка модели с использованием метрики UAR.

### Результаты исследований и их обсуждение

На первом этапе проводили эксперименты по подбору числа  $D$  извлекаемых МЧКК. Выбирался размер фрейма, равный 1024 отчета (~23 мс), классификатор на основе МОВ с линейным ядром и определялось оптимальное число  $D$  МЧКК из множества {30,32,34,36,38,40,42,44}. Наилучший результат классификации был получен при  $D = 34$ . Это значение затем использовали в дальнейших исследованиях.

На втором этапе выполнялись эксперименты с классификатором на основе МОВ и различными ядерными функциями. Известно, что функция полиномиального ядра имеет параметр  $deg$  (степень полинома), а радиальная базисная функция – параметр  $\gamma$ . Также классификатор на основе МОВ имеет параметр  $C$ , который представляет собой «бюджет» некоторой величины, отражающий степень нарушения зазора решающей границы [9]. Подобные нарушения происходят, поскольку данные в признаковом пространстве не могут быть разделены идеально. Таким образом, параметры  $deg$ ,  $\gamma$  и  $C$  являются гиперпараметрами классификатора на основе МОВ. Для выбора оптимальных значений гиперпараметров использовалась техника поиска по сетке (grid search): для  $deg$  – сетку значений от 1 до 8 с шагом 1; для  $C$  – сетку из 12 значений от 0,1 до 100,0, расположенных эквидистантно в логарифмическом масштабе; для  $\gamma$  формировалась сетка из 10 значений в диапазоне от  $10^{-3,5}$  до  $10^{-1,5}$ , расположенных эквидистантно в логарифмическом масштабе. Поиск оптимальных значений параметров, позволяющих получить наибольшее значение UAR, выполнялся для векторов признаков, полученных при различных значениях размера фрейма анализа. Результаты эксперимента приведены в табл. 1.

**Таблица 1.** Результирующий UAR для классификатора на основе метода опорных векторов с различными ядрами  
**Table 1.** Resulting UAR for a support vector machine classifier with different kernels

Размер фрейма / Frame size	Линейное ядро / Linear kernel	Полиномиальное ядро / Polynomial kernel	Радиальная базисная функция ядра / Radial basis kernel function
1024	0,458 ( $C = 0,01$ )	0,457 ( $C = 0,01, \gamma = 1, deg = 1$ )	0,469 ( $C = 8,11, \gamma = 0,0008$ )
2048	0,451 ( $C = 0,10$ )	0,450 ( $C = 0,01, \gamma = 1, deg = 1$ )	0,471 ( $C = 8,11, \gamma = 0,0008$ )
4096	0,454 ( $C = 0,01$ )	0,455 ( $C = 0,05, \gamma = 0,1, deg = 1$ )	0,476 ( $C = 2,31, \gamma = 0,0014$ )
8192	<b>0,469</b> ( $C = 0,01$ )	<b>0,474</b> ( $C = 0,05, \gamma = 0,1, deg = 1$ )	<b>0,482</b> ( $C = 28,48, \gamma = 0,0014$ )

Наилучшее значение  $UAR = 48,2\%$  достигается при использовании классификатора с ядрами РБФ и МЧКК, рассчитанными на фрейме размером 8192. Поверхность  $UAR$ , рассчитанная в ходе поиска по сетке для этой модели, представлена на рис. 6, а. Из рисунка видно, что большое значение параметра  $C$  приводит к более гибкому классификатору с высокой производительностью. При увеличении размера фрейма анализа до 16 384 показатель  $UAR$  резко падал до значения 0,13. На рис. 6, б представлена матрица спутывания для лучшей модели.

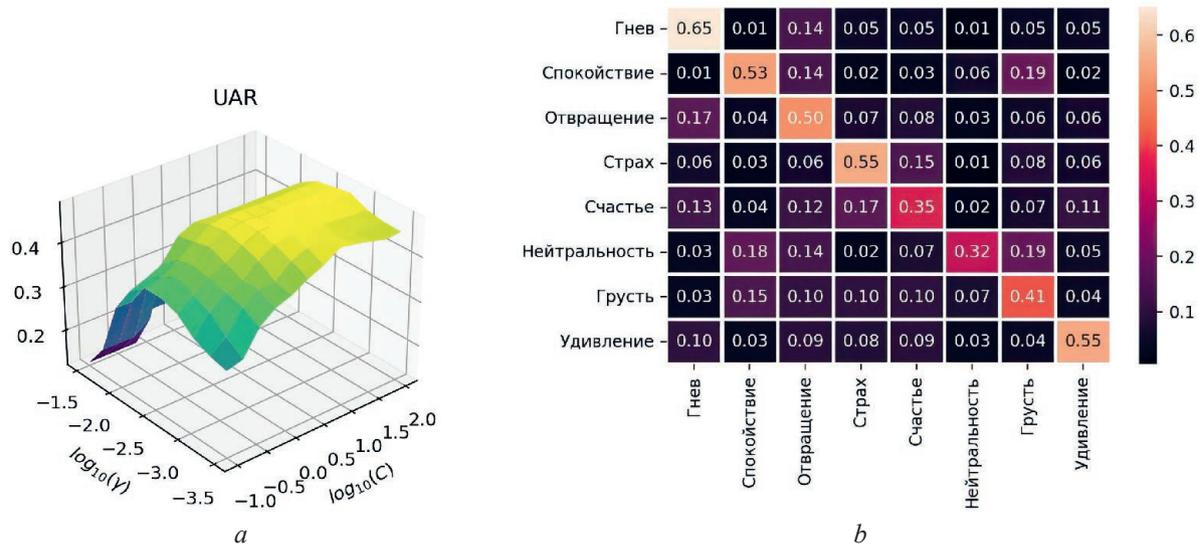


Рис. 6. Результаты экспериментов: а – поверхность UAR; б – матрица спутывания  
Fig. 6. Experimental results: a – UAR surface; b – confusion matrix

Анализ матрицы спутывания позволяет выявить важные закономерности в распознавании эмоций. Можно заметить, что наиболее часто неправильно классифицированными эмоциями являются нейтральность (32 %) и счастье (35 %). Примечательно, что нейтральность часто путается с грустью и спокойствием, что позволяет предположить некоторое сходство их акустических характеристик. И наоборот, гнев имеет высокую точность распознавания (65 %) и редко ошибочно классифицировался как другая эмоция, что указывает на отличительные особенности его акустического профиля. Эти результаты проливают свет на проблемы, с которыми сталкивается классификатор при различении тонких эмоциональных нюансов, и подчеркивают важность разработки функций и совершенствования моделей для улучшения эффективности распознавания эмоций.

Результаты исследований показывают, что выбор ядра оказывает существенное влияние на точность классификации. Классификатор с РБФ продемонстрировал высокую производительность в отношении множества эмоций, в то время как использование линейного ядра позволяет хорошо различать лишь отдельные эмоциональные состояния.

Размер фрейма, используемый для извлечения МЧКК, играет существенную роль в общей точности системы: более короткие фреймы обеспечивают более мелкие временные детали, а более длинные собирают более широкую контекстную информацию, что позволяет лучше распознавать эмоции. Полученные результаты подчеркивают важность точной настройки ядра классификатора и подбора размера фрейма при разработке систем распознавания эмоций по речи. Точность ( $UAR$ ) полученной модели, использующей 306-компонентный вектор признаков, составляет 48,2 %, что больше, чем у модели на основе МОВ [12], использующей в качестве вектора признаков 2048-компонентный вектор, полученный от внутреннего слоя нейронной сети AlexNet ( $UAR = 45,8\%$ ). Следует также отметить, что человек справляется с задачей распознавания эмоций по речи с точностью 67 % [12].

## Заключение

Рассмотрен подход к задаче распознавания эмоций по речи, основанный на классификаторе на базе метода опорных векторов и использующий надсегментные мел-частотные кепстральные коэффициенты. Наилучшие результаты ( $UAR = 48,2\%$ ) получены в случае применения клас-

сификатора с радиальной базисной функцией ядра и признаками мел-частотных кепстральных коэффициентов, рассчитанными на фрейме длительностью 170 мс.

### Список литературы / References

1. Issa D., Demirci F. M., Yazici A. (2020) Speech Emotion Recognition with Deep Convolutional Neural Networks. *Biomedical Signal Processing and Control*. 59.
2. Luna-Jiménez C., Griol D., Callejas Z., Kleinlein R., Montero J. M., Fernández-Martínez F. (2021) Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning. *Sensors*. 21 (22), 1–29.
3. Xiao H., Li W., Zeng G., Wu Y., Xue J., Zhang J., [et al.] (2022) On-Road Driver Emotion Recognition Using Facial Expression. *Appl. Sci*. 12.
4. Sadok S., Leglaive S., Séguier R. (2023) A Vector Quantized Masked Autoencoder for Speech Emotion Recognition. *arXiv preprint arXiv*. 2304.
5. Bhavan A., Chauhan P., Shah R. R. (2019) Bagged Support Vector Machines for Emotion Recognition from Speech. *Knowledge-Based Systems*. 184, 1–7.
6. Baruah M., Banerjee B. (2022) Speech Emotion Recognition via Generation Using an Attention-Based Variational Recurrent Neural Network. *Proc. Interspeech*. 4710–4714.
7. Yu C., Tian Q., Cheng F., Zhang S. (2011) Speech Emotion Recognition Using Support Vector Machines. *Advanced Research on Computer Science and Information Engineering. Communications in Computer and Information Science*. 152.
8. Huang X., Acero A., Hon H.-W., Foreword By-Reddy R. (2001) Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. *Prentice Hall PTR*.
9. Hastie T., Tibshirani R., Friedman J. H., Friedman J. H. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*.
10. On C. K., Pandiyan P. M., Yaacob S., Saudi A. (2006) Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition. *In 2006 International Conference on Computing & Informatics*. 1–5.
11. Livingstone S. R., Russo F. A. (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A Dynamic, Multimodal Set of Facial and Vocal Expressions in North American English. *PLoS One*. 13 (5).
12. Luna-Jiménez C., Griol D., Callejas Z., Kleinlein R., Montero J. M., Fernández-Martínez F. (2021) Multimodal Emotion Recognition on RAVDESS Dataset Using Transfer Learning. *Sensors*. 21.

### Вклад авторов / Authors' contribution

Авторы внесли равный вклад в написание статьи / The authors contributed equally to the writing of the article.

#### Сведения об авторах

**Краснопрошин Д. В.**, магистрант каф. электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники

**Вашкевич М. И.**, д-р техн. наук, проф. каф. электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники

#### Адрес для корреспонденции

220013, Республика Беларусь,  
г. Минск, ул. П. Бровки, 6  
Белорусский государственный университет  
информатики и радиоэлектроники  
Тел.: +375 17 293-84-78  
E-mail: vashkevich@bsuir.by  
Вашкевич Максим Иосифович

#### Information about the authors

**Krasnoproshin D. V.**, Master's Student at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics

**Vashkevich M. I.**, Dr. of Sci. (Tech.), Professor at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics

#### Address for correspondence

220013, Republic of Belarus,  
Minsk, P. Brovki St., 6  
Belarusian State University  
of Informatics and Radioelectronics  
Tel.: +375 17 293-84-78  
E-mail: vashkevich@bsuir.by  
Vashkevich Maxim Iosifovich