

DOI 10.37882/2223-2966.2026.02.10

# РАСПОЗНАВАНИЕ ЭМОЦИЙ ПО ПАРАМЕТРАМ МИМИКИ В ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

## EMOTION RECOGNITION FROM FACIAL EXPRESSION PARAMETERS IN VIRTUAL REALITY

A. Dolgikh  
E. Radostev

*Summary.* This paper presents a method for automatic recognition of a user's emotional state from facial-expression parameters (blendshape coefficients) captured by the built-in sensors of a VR headset. Emotions are represented in two ways: as one of seven basic categories and as a point in the continuous three-dimensional Valence–Arousal–Dominance (VAD) space. The proposed approach evaluates both linear models (multivariate regression and SVM) followed by mapping to the nearest emotion in the VAD space, and direct classifiers (logistic regression and a neural network) trained on discrete emotion classes. Experimental data were collected in VR and include 52 facial-expression features. Logistic regression achieved an accuracy of approximately 60 %, while the neural network reached 56 % (both demonstrating a high correlation between the predicted and ground-truth VAD coordinates), whereas linear methods yielded only 30–45 %. An ensemble classifier did not improve accuracy. In addition, we developed a neural network that generates a corresponding facial blendshape configuration for virtual avatar control given target (V, A, D) coordinates. The results are consistent with current trends (state-of-the-art systems for partially occluded faces typically report around 70 % accuracy) and support the practical feasibility of emotion recognition in VR. Typical challenges (notably confusions between closely related emotions) are discussed, and directions for integrating VR-based emotion recognition into virtual agent systems are outlined.

*Keywords:* emotion recognition, virtual reality, virtual environment, virtual avatars, machine learning, facial expressions.

**Долгих Анатолий Андреевич**

Аспирант, Национальный Исследовательский  
Ядерный Университет «МИФИ», г. Москва  
tolick.dolgih2013@yandex.ru

**Радостев Эдуард Андреевич**

Национальный Исследовательский  
Ядерный Университет «МИФИ», г. Москва  
radostevduard@gmail.com

*Аннотация.* В статье представлен метод автоматического распознавания эмоционального состояния пользователя по параметрам мимики лица (коэффициентам blendshape), считываемым встроенными датчиками VR-шлема. Эмоция описывается двумя способами: как одна из семи базовых категорий и как точка в непрерывном трёхмерном пространстве Валентность–Возбужденность–Доминантность (Valence–Arousal–Dominance, VAD). Для распознавания применены как линейные алгоритмы (многомерная регрессия, SVM) с последующим определением ближайшей эмоции в пространстве VAD, так и прямые классификаторы (логистическая регрессия, нейронная сеть) на классы эмоций. Экспериментальные данные собраны в VR и включают 52 показателя мимики лица. Логистическая регрессия достигла точности около 60 %, нейросеть 56 % (обе показали высокую корреляцию прогнозируемых координат VAD с истинными), тогда как линейные методы дали лишь 30–45 %. Ансамблевый классификатор не повысил точность. Также разработана нейросеть, которая по заданным координатам (V, A, D) генерирует соответствующую конфигурацию facial blendshape для управления виртуальным аватаром. Результаты соответствуют современным тенденциям (в современных системах для частично скрытого лица точность порядка 70 %) и подтверждают принципиальную реализуемость распознавания эмоций в VR. Отмечены типичные сложности (перекрёстные ошибки между близкими эмоциями) и обозначены перспективы интеграции VR-распознавания эмоций в системы виртуальных агентов.

*Ключевые слова:* распознавание эмоций, виртуальная реальность, виртуальное окружение, виртуальные аватары, машинное обучение, мимика.

### Введение

Виртуальная реальность (VR) широко применяется в исследованиях эмоций, поскольку позволяет моделировать реалистичные ситуации и получать выраженный эмоциональный отклик. В частности, показано, что VR может выступать эффективным инструментом эмоциональной индукции и обеспечивать более сильное эмоциональное возбуждение по сравнению с менее иммерсивными визуальными форматами [9]. Как иммерсивная технология взаимодействия, VR в прикладных (в том числе развлекательных) сценариях нацелена на создание интенсивного пользовательского

опыта и «эффекта присутствия», что связывают с необходимостью глубокого погружения в виртуальную среду [2]. Одновременно VR-платформы требуют надёжного автоматического распознавания эмоционального состояния пользователя по данным, собираемым в процессе взаимодействия, причём современная литература подчёркивает важность сочетания иммерсивных сценариев с неинвазивными измерениями и контролем стимулов [8]. Автоматическое определение эмоций по таким сигналам важно для реалистичных социальных взаимодействий в VR, где требуется правдоподобная передача выражений эмоций у виртуальных представлений пользователей [2].

Распознавание эмоций по выражению лица является ключевым компонентом систем человеко-компьютерного взаимодействия. Однако в условиях VR эта задача осложнена: шлем виртуальной реальности закрывает значительную часть лица пользователя, затрудняя регистрацию мимики стандартными методами компьютерного зрения [5]. Для социального VR это критично, поэтому используются способы повысить выразительность и точность передачи эмоций через аватар, включая решения с трекингом лица и взгляда [5][6]. Один из подходов опирается на встроенные в гарнитуру камеры слежения за глазами: анализ изображений области глаз позволяет классифицировать часть выражений без внешней камеры [5]. Альтернативный подход — использование дополнительных сенсоров, регистрирующих физиологические реакции, и применение этих данных для распознавания эмоций в VR [8]. Появление подобных технологий открывает возможность интеграции модуля распознавания эмоций непосредственно в VR.

Для представления эмоций в статье используется трёхмерная модель «валентность–возбужденность–доминантность» (Valence–Arousal–Dominance, VAD). Для сопоставления дискретных классов и непрерывных координат были использованы фиксированные численные значения VAD для шести базовых эмоций и нейтрального состояния.

Таким образом, цель настоящего исследования — разработать и экспериментально оценить подход к распознаванию эмоций пользователя в VR-среде по параметрам мимики лица, а также изучить возможности применения такого решения в системах виртуальной реальности.

### Литературный обзор

Влияние VR-технологий на эмоциональные состояния активно изучается в последнее десятилетие. Показано, что виртуальная среда способна усиливать эмоциональные реакции пользователя по сравнению с неиммерсивными методами предъявления стимулов [7, 9]. Благодаря эффекту присутствия применение VR в исследованиях и прикладных задачах позволяет вызывать более сильные эмоции и тем самым усиливать эмоциональное воздействие [8]. Вместе с тем эмоционально-насыщенные VR-сценарии требуют надёжного мониторинга состояния пользователя. Возникает необходимость автоматического распознавания эмоций по данным, собираемым в VR (мимике лица, поведению, физиологическим показателям и т.д.), что критично для реализации реалистичных социальных взаимодействий в виртуальной среде. При этом в современных работах всё чаще используется не только дискретное распознавание «названий» эмоций, но и их представление в непрерывном двух- или трёхмерном пространстве, поскольку оно

лучше подходит для описания сложных и смешанных аффективных состояний. В частности, модель valence–arousal–dominance (VAD) рассматривается как удобный 3D-каркас для представления эмоций и последующего распознавания по измеряемым сигналам [10].

Классические системы распознавания эмоций по выражению лица активно развиваются и в современных работах, как правило, опираются на методы глубокого обучения [5, 7]. Однако в VR значительная часть лица скрыта гарнитурой (HMD), что затрудняет анализ мимики стандартными методами компьютерного зрения [5, 7]. Для компенсации этого предложено несколько подходов. Один из них — анализ доступной области лица изнутри шлема. Например, инфракрасная камера (камера трекинга взгляда), встроенная в VR-гарнитуру, способна фиксировать движения вокруг глаз, достаточные для распознавания по крайней мере пяти выражений лица со средней точностью около 74 %, не ниже человеческой (в эксперименте — выше оценок наблюдателей) [5]. Другой подход — оснащение шлемов специальными модулями отслеживания мимики. Современные устройства (Meta Quest Pro и др.) имеют встроенные сенсоры/внутренние камеры, позволяющие получать параметры выражений лица и оперативно переносить их на модель аватара [11]. Недавнее исследование на шлеме Quest Pro с использованием облегчённой нейросети (MobileNet V2) показало высокую точность распознавания некоторых эмоций (например, радость, грусть, удивление), тогда как «гнев» и «страх» часто путались с нейтральным состоянием — эти ошибки авторы связывают с ограничениями текущих моделей и технологии распознавания [11]. Другая работа представила специализированный датасет изображений лиц в VR (с частично скрытым лицом) и обучила свёрточный классификатор EfficientNet, который достиг около 70 % точности по семи классам эмоций [7]. Это заметно ниже показателей для открытого лица, но подтверждает принципиальную реализуемость классификации выражений в VR при некотором снижении качества. В целом в условиях естественного перекрытия лица гарнитурой (шлемом) в экспериментальных работах демонстрируется порядка 70 % точности по семи классам эмоций, что уступает классическим условиям, но свидетельствует о преодолении ключевых технических барьеров [7].

Ещё одно перспективное направление — одновременное использование нескольких источников данных. Так, в работе Bastida и соавт. (2024) одновременно анализировались мимические и физиологические сигналы пользователя в VR; для объединения результатов классификаторов применена схема позднего объединения (late fusion) с учётом того, что распознавание по лицу затрудняется VR-оборудованием [3]. В другом исследовании эмоции пользователей вызывались VR-сценариями и распознавались по физиологическим сигналам: достиг-

нута порядка 70 % точности для 4-классовой классификации эмоций и до 80–85 % при бинарной классификации высокой/низкой валентности или активации [8]. Эти результаты подтверждают, что VR-технологии, несмотря на создаваемые неудобства с гарнитурой, предоставляют новые возможности — они легко сочетаются с носимыми сенсорами, позволяя собирать богатый материал для анализа эмоций [3][8]. Современные исследования показывают, что объединение нескольких источников сигналов может повышать надёжность распознавания эмоций по сравнению с опорой на одну модальность [3].

Одно из ключевых направлений практического применения распознавания эмоций — создание эмоциональных виртуальных собеседников и помощников. В научных исследованиях описаны прототипы голосовых ассистентов, которые с помощью нейросетевых моделей определяют эмоциональную окраску речи пользователя и адаптируют стиль ответа для более эмпатичного диалога [12]. Другая важная сфера — виртуальные персонажи и аватары в VR/AR-системах и играх. Распознавание эмоций пользователя может использоваться для адаптации VR-контента/параметров среды под текущее эмоциональное состояние, повышая персонализацию опыта [6]. Текущие VR-технологии (в частности, facial blendshape) уже позволяют транслировать мимику пользователя на модель лица аватара. Если же требуется определять именно название эмоции, а не сырые параметры выражения, связка из модуля распознавания и последующей генерации соответствующей анимации лица оказывается востребованной. Концептуально близкая идея обсуждается в работе Зиннатова и Кугураковой (2020): авторы рассматривают объединение управляемой лицевой анимации с модулями учета эмоционального состояния и микромимики. В их реализации по аудиопотоку (в т.ч. записанной или синтезированной речи) автоматически строится анимация 3D-персонажа, а учет эмоций предлагается расширять за счёт дополнительного анализа текста/аудио и других источников, что ориентировано на сценарии, где текст и речь заранее неизвестны [1]. Интеграция модулей распознавания и генерации эмоций в целом повышает реалистичность и выразительность виртуальных агентов, приближая их к человеческим стандартам общения.

## Материалы и методы

### Датасет и разметка

Для обучения и оценки моделей был собран собственный набор данных эмоциональных состояний пользователей в виртуальной реальности. В эксперименте участвовали 15 добровольцев (совершеннолетние мужчины и женщины), каждый взаимодействовал с несколькими VR-сценариями, вызывающими разные эмоции.

В качестве визуальных стимулов использовались изображения из набора Karolinska Directed Emotional Faces (KDEF), представляющего стандартизированные выражения для базовых эмоций. Для эксперимента был сформирован поднабор из 56 изображений: по 4 мужских и 4 женских лиц для каждой из семи эмоций (включая нейтральное выражение). Такой отбор обеспечивает баланс по полу и по классам эмоций при предъявлении стимулов.

Сбор данных проводился в VR-сцене, включающей (1) режим демонстрации слайдов со стимулами и (2) режим «зеркала» для самоконтроля выражения. На первом этапе испытуемому предъявлялись изображения лицевых выражений из KDEF, и предлагалось воспроизвести выражение; при этом выполнялась запись эпизода и автоматически фиксировалась метка эмоции, соответствующая предъявленному стимулу. На втором этапе визуальный стимул отсутствовал: испытуемому называлась эмоция, которую он воспроизводил, ориентируясь на отражение своего аватара, а запись выполнялась тем же способом. Для предъявления использовался набор из 56 изображений-стимулов из KDEF.

Испытуемый, видя стимул, переживал или изображал соответствующую эмоцию в VR. Пользователь носил шлем HTC Vive Pro Eye с модулем Vive Facial Tracker (инфракрасная камера для регистрации мимики нижней части лица). Аппаратно в реальном времени регистрировались:

- Параметры мимики лица. Трекинг-модуль фиксировал мимику ключевых областей лица пользователя, представляя её в виде набора из 52 коэффициентов facial blendshape. Каждый коэффициент отражает степень выраженности определённого элементарного движения лица (поднятие брови, улыбка, прищуривание и т.д.).
- Прочие сигналы. Дополнительно собирались данные движения глаз (направление взгляда, частота моргания).

Сырые данные мимики сохранялись в формате CSV. Таблица содержала временные метки, после чего следовали значения 52 blendshape. Такой формат упрощает синхронизацию с внешними сигналами и последующую обработку.

В ходе сбора данных использовался режим фиксации, при котором запись начинается по нажатию управляющей кнопки и завершается вторым нажатием; итоговые значения признаков рассчитываются усреднением по всем кадрам, накопленным за время эпизода. Такой подход снижает влияние моментных артефактов трекинга и позволяет представлять каждое наблюдение единым вектором из 52 коэффициентов (параметры нижней и верхней частей лица), соответствующим устойчиво

удерживаемому выражению. Помимо данного режима, предусмотрены режим «снимка» и режим усреднения по фиксированному числу кадров, однако в настоящей работе при сборе датасета применялся именно режим записи «по двум нажатиям» с усреднением по всем кадрам эпизода.

Каждый записанный эпизод снабжался категориальной меткой одной из семи эмоций (гнев, радость, удивление, отвращение, испуг, грусть, нейтральность). Метка определялась процедурой предъявления: на этапе со слайдами-стимулами она задавалась подписью к стимулу, а на этапе без визуального стимула — называлась экспериментатором и воспроизводилась испытуемым. Таким образом, для каждого эпизода известен класс эмоции, использованный далее, как целевая переменная для прямых классификаторов. Для моделей, работающих в непрерывном пространстве VAD, категориальная метка дополнительно преобразовывалась в координаты (V,A,D) по табличному соответствию базовых эмоций (таблица 1), после чего решалась задача регрессии по каждому измерению VAD с последующей проекцией предсказанной точки на ближайший эмоциональный класс. Нейтральное состояние в работе задавалось как (0,0,0) (центр шкалы).

Таблица 1.

Значения 6-основных эмоций в терминах VAD [4]

	Валентность	Возбуждённость	Доминантность
Гнев	-0,43	0,67	0,34
Радость	0,76	0,48	0,35
Удивление	0,4	0,67	-0,13
Отвращение	-0,6	0,35	0,11
Испуг	-0,64	0,6	-0,43
Грусть	-0,63	0,27	-0,33

В рамках первичного сбора данных было получено 497 записей от 15 участников; все записи собирались в единый CSV и использовались для обучения сравниваемых моделей. Данные случайным образом разделялись на обучающую и валидационную выборки.

#### Модели распознавания

Для автоматической классификации эмоций по данным мимики были реализованы несколько алгоритмов машинного обучения. Линейные модели — многомерная линейная регрессия и метод опорных векторов (SVM) — обучались предсказывать непрерывные координаты эмоционального состояния в пространстве VAD; затем полученная точка (V, A, D) проецировалась на ближайший класс эмоций (выбиралась минимальная дистанция до эталонных координат 7 базовых эмоций). Помимо

этого, использовались прямые методы классификации на 7 категорий: логистическая регрессия и полносвязная нейронная сеть. Архитектура нейросети включала три скрытых слоя по 64 нейрона (с функцией Leaky ReLU) и выходной слой из 7 нейронов с softmax-активацией. Обучение проводилось на собранном наборе данных, используя категориальные метки эмоций.

Проверялась также эффективность ансамблей моделей. Были реализованы группы из трёх логистических регрессий и из трёх нейронных сетей; решением ансамбля считался эмоциональный класс, предсказанный большинством моделей.

Отдельно была обучена и обратная модель на нейросетевой основе, выполняющая задачу генерации выражения лица по заданным эмоциональным координатам. Эта модель принимала на вход желаемое значение валентности, активации и доминантности (V, A, D) и выдавала 52 коэффициента facial blendshape, которые можно применить к аватару. Таким образом, реализована возможность синтезировать мимику виртуального персонажа под заданную эмоцию.

#### Метрики оценки

Качество моделей оценивалось по доле правильных классификаций (accuracy) на отложенной выборке. Дополнительно для подходов, использующих пространство VAD, вычислялись корреляции между предсказанными и эталонными значениями валентности, активации и доминантности. Для классификаторов значения VAD восстанавливались через сопоставление каждому классу его координат (V, A, D) по таблице 1. Аналогично формировались «эталонные» значения VAD из истинных категориальных меток. При запуске обучающих скриптов все эпизоды случайным образом разделялись на обучающую и валидационную выборки; в отдельных экспериментах при необходимости дополнительно формировалась тестовая выборка, однако основная оценка качества в работе приводится по валидационному набору.

Поскольку датасет содержит множественные эпизоды от каждого участника, случайное разбиение по эпизодам потенциально допускает попадание записей одного и того же человека в обе выборки, что может завышать оценку обобщающей способности на новых пользователей. В дальнейшем планируется оценка по протоколу разбиения по участникам, чтобы измерять переносимость модели на ранее не встречавшихся людей.

#### Результаты

В таблице 2 продемонстрированы результаты работы модели линейной регрессии и метода опорных век-

торов, с использованием евклидова и манхэттенского расстояний. Также в таблице указана общая точность данных моделей.

Несмотря на то, что обе модели линейные, метод опорных векторов показывает заметно лучший результат на обеих метриках расстояния. Тем не менее, результат можно трактовать как предсказание нейтральности и отвращения, а иногда радости и удивления. Злость, грусть и страх предсказывается во всех четырех вариантах достаточно плохо.

Таблица 2.

Результаты работы линейных моделей

Модель	Линейная регрессия		Метод опорных векторов	
	Эвклид	Манхэттен	Эвклид	Манхэттен
Расстояние				
Удивление	22 %	36.5 %	51.1 %	51.1 %
Радость	51 %	51 %	58.4 %	58.4 %
Грусть	14.6 %	14.6 %	29.2 %	21.9 %
Страх	0	0	29.2 %	29.2 %
Злость	0	0	0	0
Отвращение	43.8 %	43.8 %	58.4 %	73 %
Нейтральность	73 %	73 %	80.3 %	80.3 %
Общая точность	29.2 %	31.25 %	43.75 %	44.8 %

Модель логистической регрессии продемонстрировала общую точность порядка 60 %. В таблице 3 приведены коэффициенты корреляции между предсказанными и эталонными значениями VAD, вычисленными на основе выходных эмоциональных оценок модели. Анализируя результат работы логистической модели, можно увидеть заметную корреляцию получаемых значений VAD, а также достаточно высокую общую точность, в сравнении с линейными моделями. Матрица ошибок для логистической регрессии приведена на рисунке 1-а.

Таблица 3.

Корреляция значений VAD предсказанных и истинных эмоций для логистической модели

Общая точность: 60 %	Предсказанные значения			
	Valence	Arousal	Dominance	
Истинные значения	Valence	0.63814	–	–
	Arousal	–	0.67871	–
	Dominance	–	–	0.674081

Далее, в таблице 4 приведены коэффициенты корреляции для значений Valence, Arousal и Dominance, восстановленных на основе эмоций, предсказанных нейросетевой моделью.

Таблица 4.

Корреляция значений VAD предсказанных и истинных эмоций для нейросетевой модели

Общая точность: 56.25 %	Предсказанные значения			
	Valence	Arousal	Dominance	
Истинные значения	Valence	0.516776	–	–
	Arousal	–	0.43347	–
	Dominance	–	–	0.671306

Приведенные в таблице 4 значения работы нейросетевой модели демонстрируют заметную корреляцию получаемых значений валентности–возбужденности–доминантности (VAD), а также достаточно высокую общую точность определения эмоций в сравнении с линейными моделями (таблица 2). Матрица ошибок для данного подхода приведена на рисунке 1-б.

Результаты работы ансамбля из 3 моделей логистической регрессии на валидационном наборе данных оказались несколько хуже обычной логистической регрессии. Общая точность при использовании ансамбля составила 47 %. Матрица ошибок приведена ниже на рисунке 2-а.

На рисунке 1-б приведена матрица ошибок, полученная в результате работы ансамбля из 3 нейросетевых моделей на валидационном наборе данных. Общая точность при использовании данного метода составила 54 %, при том, что точность каждой из входящих моделей на несколько процентов выше.

По приведенным выше матрицам ошибок можно проследить, что одни эмоции распознаются значительно лучше, чем другие. Можно заметить, что модели часто путают испуг и удивление. Также обращает на себя внимание то факт, что столбец с нейтральной эмоцией ярче остальных для всех матриц ошибок. Доминирование нейтрального класса в матрицах ошибок согласуется с наблюдением о слабой экспрессии части испытуемых: при недостаточно выраженной мимике модели склонны «усреднять» предсказание и относить эпизод к нейтральному состоянию. Это также указывает на необходимость расширения выборки и усиления вариативности выражений, поскольку текущий объем данных может быть недостаточен для устойчивого обучения, особенно в постановке «классификация эмоций», тогда как для задач в терминах VAD более естественной является регрессионная постановка.

Применение ансамблевых моделей не принесло значительного прироста качества. Ансамбль из трёх логистических регрессий показал 47 % точности (хуже, чем одна модель), а ансамбль трёх нейросетей — 54 % (на уровне одиночной сети). Причиной этого стало, вероятно, сильное пересечение ошибок между моделью

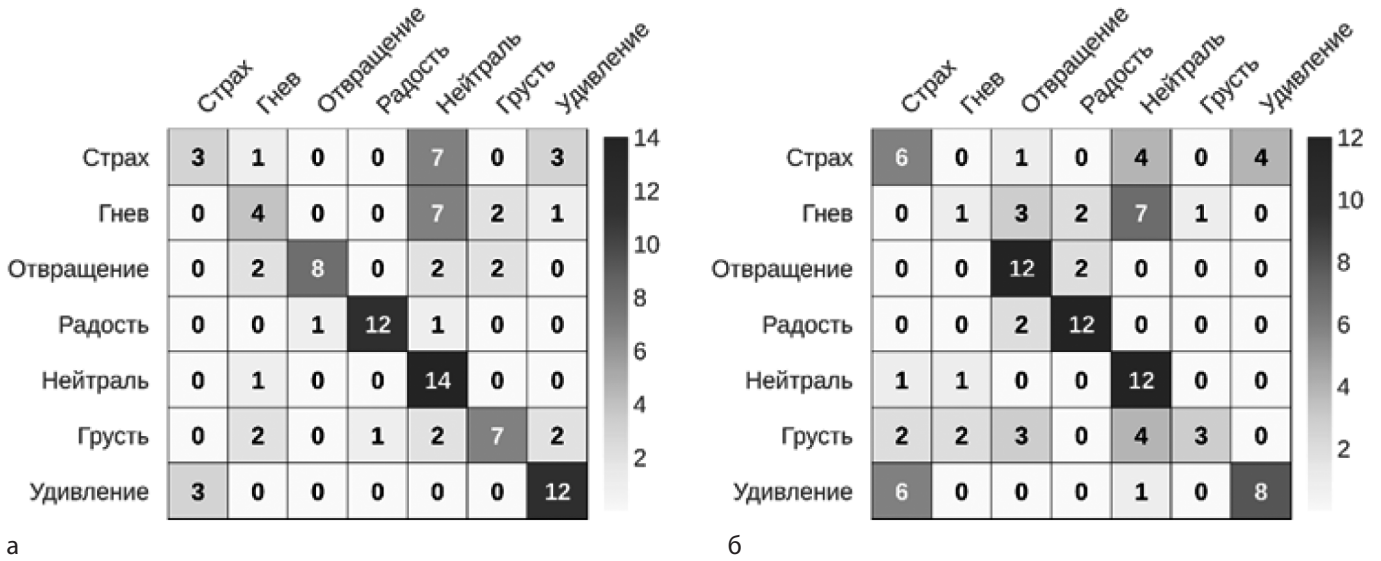


Рис. 1. Матрицы ошибок (Confusion Matrix)  
 а) матрица ошибок для логистической регрессии; б) матрица ошибок нейросетевой модели

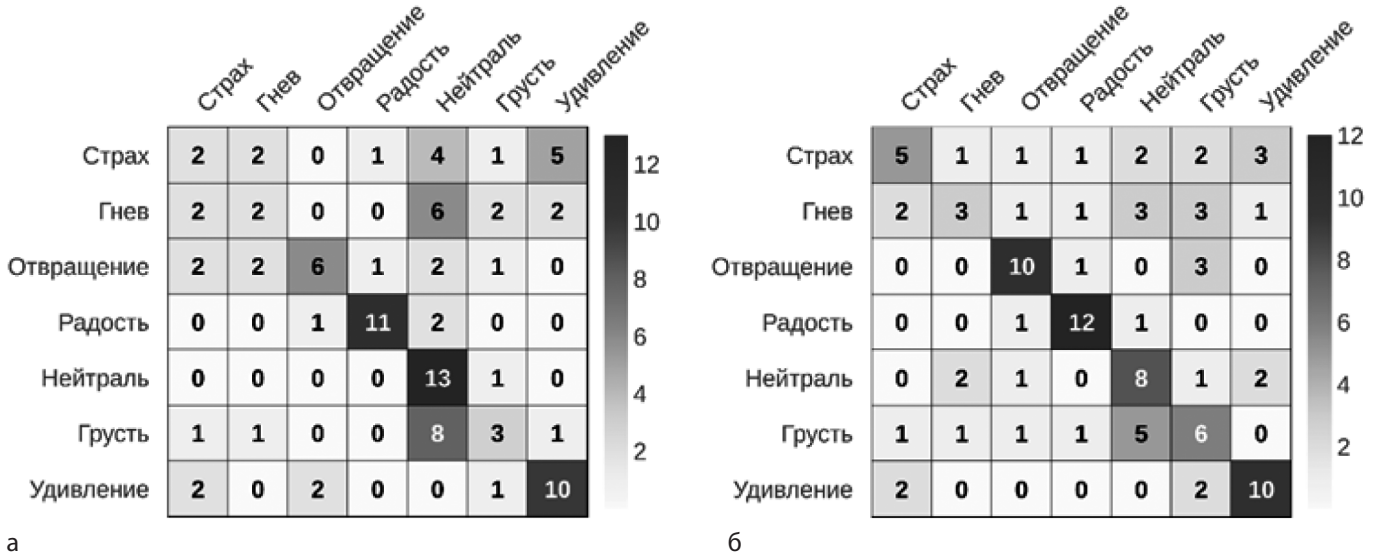


Рис. 2. Матрицы ошибок (Confusion Matrix) для ансамблевых моделей

а) матрица ошибок ансамбля моделей логистической регрессии; б) матрица ошибок ансамбля нейросетевых моделей

ми: одинаковые по архитектуре классификаторы, обученные на одном и том же наборе, склонны ошибаться на одних и тех же примерах, поэтому усреднение не помогает. Кроме того, правило большинства могло приводить к проигрышу в ситуациях, когда одна из моделей была права, а две — одинаково ошибались. Таким образом, одиночная логистическая регрессия оказалась не только самой простой, но и одной из самых результативных моделей на наших данных. Её преимущество может объясняться меньшей склонностью к переобучению по сравнению с более сложной нейронной сетью при ограниченном размере выборки.

### Обсуждение

Полученные результаты отражают как общие тенденции автоматического анализа эмоций, так и специфиче-

ские нюансы VR-контекста. Во-первых, как и в традиционных системах распознавания по лицу, одни эмоции распознаются значительно легче, чем другие. Нейтральное выражение определяется моделью с наибольшей надёжностью, тогда как некоторые эмоциональные состояния (особенно «гнев» и «страх») оказываются трудными для автоматического различения. Проведенные эксперименты демонстрируют аналогичные сложности: негативные эмоции классифицировались существенно хуже, чем позитивные, и часто путались между собой или с нейтральным состоянием. Подобные проблемы отмечаются и в других исследованиях: например, даже продвинутые нейросетевые классификаторы иногда испытывают сложности с разграничением схожих эмоций. Причины ошибок в нашем случае связаны как с неполнотой входных данных (VR-шлем скрывает часть лица, ли-

шая алгоритм важных признаков), так и с ограниченным объёмом обучающей выборки. Для преодоления этих ограничений необходим либо учёт дополнительных модальностей (речь, физиология и др.), либо использование более сложных архитектур, способных улавливать тонкие различия выражений.

Если сопоставить достигнутую точность с результатами других разработок, можно сказать, что она лежит в том же порядке величины: порядка 60–70 % для распознавания эмоций по лицу в VR. Например, алгоритм Eyemotion на основе трекинга глаз достигал 74 % (для ограниченного набора выражений), а свёрточная модель на частично скрытых лицах — около 70 % на 7 эмоциях. Это свидетельствует о том, что качество автоматического определения 6–7 эмоций в VR в настоящее время находится на среднем уровне и существенно уступает традиционному анализу лица без помех. Тем не менее наши эксперименты подтверждают принципиальную осуществимость распознавания эмоций даже при наличии гарнитуры, а предложенный подход отличается низкой вычислительной сложностью и непосредственной работой с сигналами VR-шлема. Последнее особенно важно для внедрения в интерактивные приложения: логистическая модель выполняется за доли миллисекунды на CPU, что позволяет запускать её в реальном времени даже на мобильных VR-устройствах.

### Заключение

В работе разработан и экспериментально оценён подход к автоматическому распознаванию эмоционального состояния пользователя в виртуальной реальности по данным трекинга мимики (52 коэффициента facial blendshape), регистрируемым VR-гарнитурой. Рассмотрены два представления эмоций: дискретная классификация по 7 базовым классам и непрерывное описание

в пространстве Valence–Arousal–Dominance (VAD), используемое как для анализа результатов, так и для последующего управления выражением лица аватара.

Сравнение моделей показало, что наилучшее качество на собранных данных обеспечивает логистическая регрессия (accuracy около 60 %), тогда как полносвязная нейросеть с тремя скрытыми слоями демонстрирует сопоставимый результат (около 56 %). Линейные подходы, предсказывающие координаты VAD с последующей проекцией на ближайший класс, уступают по точности (порядка 30–45 %), что указывает на преимущественно нелинейную зависимость между параметрами мимики и воспринимаемой эмоцией. Ансамбли однородных моделей не дали прироста качества, что объясняется схожестью их ошибок. Представление VAD подтвердило свою практическую полезность: корреляции предсказанных и эталонных компонент достигали примерно 0,6–0,7 и позволяли интерпретировать близость эмоций и характер ошибок. Дополнительно реализована обратная модель, генерирующая конфигурацию blendshape по заданным координатам (V, A, D), что формирует основу связки «распознавание + генерация» для управления мимикой виртуального персонажа в реальном времени.

Наибольшие затруднения связаны с различием визуально близких эмоций (в частности, перекрёстные ошибки между испугом и удивлением, а также преобладание нейтрального класса при низкой уверенности), что обусловлено частичной окклюзией лица гарнитурой и ограниченным объёмом выборки. В дальнейших исследованиях целесообразно расширить и разнообразить данные, перейти к протоколам оценки с разбиением по участникам, а также развивать мультимодальные решения (мимика + речь + физиологические сигналы) и методы переноса обучения, компенсирующие неполную наблюдаемость лица в VR.

### ЛИТЕРАТУРА

1. Зиннатов А.А., Кугуракова В.В. Механизмы реалистичной мимики для антропоморфных социальных агентов // Электронные библиотеки. — 2020. — Т. 23, № 5. — С. 1011–1025. DOI: 10.26907/1562-5419-2020-23-5-1011-1025
2. Кармадонов В.Ю. Совмещение технологий отслеживания взгляда, распознавания эмоций и VR // Academy. — 2018. — № 7 (34). — С. 41–43.
3. Bastida L., Sillaurren S., Loizaga E., Tomé E., and Moya A. Exploring Human Emotions: A Virtual Reality-Based Experimental Approach Integrating Physiological and Facial Analysis // Multimodal Technologies and Interaction. — 2024. — Т. 8. — No. 6. — С. 47. DOI: 10.3390/mti8060047
4. Buechel S., Hahn U. Emotion Analysis as a Regression Problem — Dimensional Models and Their Implications on Emotion Representation and Metrical Evaluation // 22nd European Conference on Artificial Intelligence — IOS Press, 2016. — Vol. 285, P. 1114–1122. DOI: 10.3233/978-1-61499-672-9-1114
5. Hickson S., Dufour N., Sud A., Kwatra V., Essa I. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras // 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). — IEEE, 2019. — P. 1626–1635. DOI: 10.1109/WACV.2019.00178
6. Linares-Vargas B.G.P., Cieza-Mostacero S.E. Interactive virtual reality environments and emotions: A systematic review // Virtual Reality. — 2024. — Т. 29. — No. 1. — С. 3. DOI: 10.1007/s10055-024-01049-1
7. Ortmann T., Wang Q., Putzar L. EmojiHeroVR: A study on facial expression recognition under partial occlusion from head-mounted displays // 2024 12th International Conference on Affective Computing and Intelligent Interaction (ACII). — IEEE, 2024. — С. 80–88. DOI: 10.1109/ACII63134.2024.00014.
8. Polo E.M., Iacomì F., Rey A.V., Ferraris D., Paglialonga A., Barbieri R. Advancing emotion recognition with Virtual Reality: A multimodal approach using physiological signals and machine learning // Computers in Biology and Medicine. — 2025. — Т. 193. — С. 110310. DOI: 10.1016/j.combiomed.2025.110310
9. Tian F., Hua M., Zhang W., Li Y., Yang X. Emotional arousal in 2D versus 3D virtual reality environments // PLOS ONE. — 2021. — Т. 16. — No. 9. — С. e0256211. DOI: 10.1371/journal.pone.0256211

10. Verma G.K., Tiwary U.S. Affect representation and recognition in 3D continuous valence–arousal–dominance space // *Multimedia Tools and Applications*. — 2017. — Т. 76. — No. 2. — С. 2159–2183. DOI: 10.1007/s11042-015-3119-y
11. Zhang Z., Fort J.M., Giménez Mateu L. Facial expression recognition in virtual reality environments: challenges and opportunities // *Frontiers in Psychology*. — 2023. — Т. 14. — С. 1280136. DOI: 10.3389/fpsyg.2023.1280136
12. Ma Y., Zhang Y., Bachinski M., Fjeld M. Emotion-aware voice assistants: Design, implementation, and preliminary insights // *Proceedings of the Eleventh International Symposium of Chinese CHI*. — 2023. — С. 527–532. DOI: 10.1145/3629606.3629665

---

© Долгих Анатолий Андреевич (tolick.dolgih2013@yandex.ru); Радостев Эдуард Андреевич (radosteveduard@gmail.com)  
Журнал «Современная наука: актуальные проблемы теории и практики»