

УДК 004.415.53

***СИСТЕМА РАСПОЗНАВАНИЯ ЖЕСТОВОЙ РЕЧИ  
В РЕАЛЬНОМ ВРЕМЕНИ ДЛЯ ПОДДЕРЖКИ  
ЛЮДЕЙ С НАРУШЕНИЯМИ СЛУХА***

***Сафина Г.Ф.***

*к. ф.-м. н, доцент,*

*ФГБОУ ВО «Уфимский университет науки и технологий», Нефтекамский филиал,*

*Нефтекамск, Россия*

***Маньянов А.Р.***

*студент,*

*ФГБОУ ВО «Уфимский университет науки и технологий», Нефтекамский филиал,*

*Нефтекамск, Россия*

**Аннотация:** В данной статье представлен прототип системы распознавания жестовой речи в реальном времени, ориентированной на поддержку людей с нарушениями слуха. Описывается архитектура программного решения, реализованного на языке Python с использованием библиотек OpenCV, MediaPipe и NumPy. Классификация жестов выполняется гибридным алгоритмом, сочетающим правилую логику (для основных жестов) и метод k ближайших соседей (KNN) с параметром  $k=5$ . Для обучения и тестирования собран датасет из 432 размеченных образцов, включающий 31 класс жестов (буквы русского алфавита и управляющие команды «ЛАЙК», «ДИЗЛАЙК»). На тестовой выборке (86 образцов) система достигла точности 75,6% при частоте обработки 24 кадра в секунду.

**Ключевые слова:** жестовый язык, распознавание жестов, компьютерное зрение, машинное обучение, MediaPipe, OpenCV, KNN, гибридная классификация, люди с нарушениями слуха.

***REAL-TIME SIGN LANGUAGE GESTURE RECOGNITION SYSTEM  
FOR SUPPORTING PEOPLE WITH HEARING IMPAIRMENTS***

***Safina G.F.***

*PhD, Associate Professor,*

*Neftekamsk branch of the Ufa University of Science and Technology,*

*Neftekamsk, Russia*

***Manyanov A.R.***

*student,*

*Neftekamsk branch of the Ufa University of Science and Technology,*

*Neftekamsk, Russia*

**Abstract:** This article presents a prototype of a real-time gesture speech recognition system aimed at supporting people with hearing impairments. The architecture of a software solution implemented in Python using the OpenCV, MediaPipe and NumPy libraries is described. Gesture classification is performed by a hybrid algorithm combining rule logic (for basic gestures) and the k nearest neighbors (KNN) method with parameter  $k=5$ . For training and testing, a dataset of 432 labeled samples has been assembled, including 31 classes of gestures (letters of the Russian alphabet and the control commands "LIKE" and "DISLIKE"). In the test sample (86 samples), the system achieved an accuracy of 75.6% at a processing rate of 24 frames per second.

**Keywords:** sign language, gesture recognition, computer vision, machine learning, MediaPipe, OpenCV, KNN, hybrid classification, people with hearing impairments.

В условиях цифровизации общества и активного развития сервисов, ориентированных на человека, особую значимость приобретают технологии, повышающие доступность среды для людей с ограниченными возможностями здоровья. Люди с нарушениями слуха и речи часто сталкиваются с трудностями при обращении в учреждения, взаимодействии с организациями и в повседневном общении, поскольку большинство слышащих не владеют

Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ ЭЛ № ФС 77-68405 ISSN 2541-8327

жестовым языком. Не во всех ситуациях возможно привлечение профессионального сурдопереводчика, что создаёт дополнительные барьеры.

Одним из перспективных направлений снижения подобных барьеров является разработка автоматизированных систем распознавания жестовой речи в реальном времени.

Такие системы способны преобразовывать выполняемые пользователем жесты в текст или звук, что потенциально облегчает коммуникацию между глухими и слышащими людьми в бытовых, образовательных и сервисных сценариях [1, 4].

Технической основой для создания подобных систем являются методы компьютерного зрения и машинного обучения. Использование современных библиотек, таких как OpenCV и MediaPipe, позволяет обрабатывать видеопоток с обычной веб-камеры, выделять ключевые точки кистей рук и формировать числовое представление жестов, пригодное для последующей классификации [3, 6].

В данной работе в качестве языка программирования выбран Python, поскольку он предоставляет широкий набор библиотек для анализа данных, обработки изображений и машинного обучения. Для работы с видеопотоком применяется библиотека OpenCV, обеспечивающая захват кадра с веб-камеры, преобразование цветовых пространств и базовые операции предобработки [6].

Для извлечения информативных признаков из изображения рук используется фреймворк MediaPipe, который предоставляет заранее обученные модели для детектирования кистей и определения координат 21 ключевой точки (landmarks). Модуль MediaPipe Hands позволяет получать для каждой руки набор трёхмерных координат основных суставов пальцев и ладони. Такие координаты формируют основу признакового вектора, который далее обрабатывается алгоритмами классификации [4, 6].

Архитектура разработанного прототипа опирается на модульный подход. Общая схема взаимодействия модулей представлена на рисунке 1.

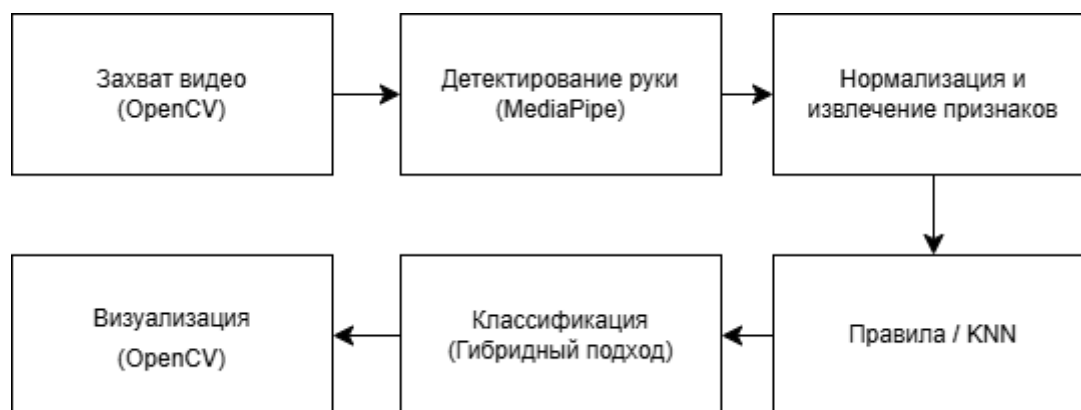


Рисунок 1 – Архитектура системы (авторская разработка)

Модуль захвата видеопотока, реализованный с использованием OpenCV, инициализирует подключения к веб-камере и читает. Модуль предобработки при необходимости выполняет изменение размера изображения и преобразование цветового пространства.

Ключевой компонент – модуль извлечения признаков – применяет к кадру модель MediaPipe Hands, возвращающую координаты ключевых точек. Полученные координаты нормализуются относительно запястья и масштабируются, что обеспечивает инвариантность к положению руки в кадре и расстоянию до камеры.

Для классификации жестов реализован гибридный подход. Сначала проверяются жесткие правила, определенные на основе конфигурации пальцев (битового вектора из 5 признаков). Например, если все пальцы сжаты в кулак – система возвращает «А», если все пальцы разжаты и ладонь повернута к камере – «В». Правила покрывают около 30 жестов и работают мгновенно. Если правило не сработало, подключается метод k ближайших соседей (KNN) с параметром  $k=5$ , который анализирует 63-мерное пространство признаков (21 точка  $\times$  3 координаты) и находит ближайшие образцы из датасета [7].

MediaPipe Hand Landmarker возвращает для каждой руки 21 ключевую точку (рисунок 2).

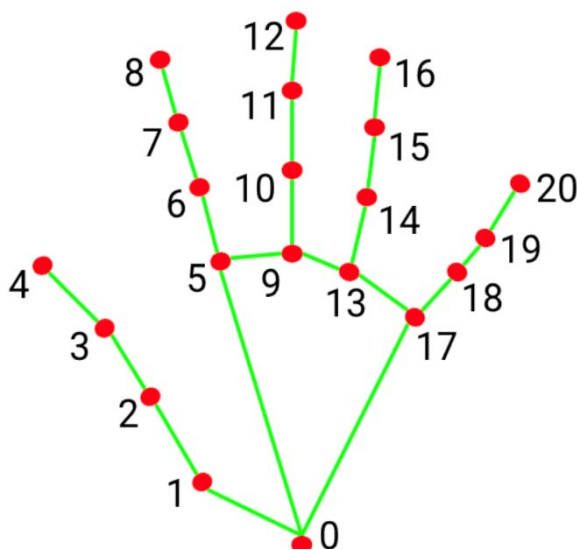


Рисунок 2 – Используемые точки руки (авторская разработка)

Особую сложность для распознавания представляют паражестные пары – жесты со схожей конфигурацией пальцев, отличающиеся незначительно. В русской жестовой азбуке к таким парам относятся М/Т и П/Л.

Для их различения в системе используются пороговые значения, определённые экспериментально. Для пары М/Т вычисляется расстояние между кончиками указательного и среднего пальцев, нормализованное к размеру ладони: если значение меньше 0,18 – жест «Т» (пальцы сомкнуты), если больше 0,22 – жест «М» (пальцы разведены). Для пары П/Л пороги составляют 0,24 (сомкнутые пальцы → «П») и 0,30 (разведённые → «Л»).

Для обучения и тестирования системы был собран собственный датасет. Всего собрано 434 изображения, успешно извлечено 432 образца (99,5%). Датасет содержит 31 класс жестов: буквы русского алфавита и управляющие жесты «ЛАЙК» и «ДИЗЛАЙК». Разделение на обучающую и тестовую выборки выполнено в соотношении 80/20: обучающая выборка – 346 образцов, тестовая – 86 образцов.

Результат распознавания отображается пользователю в виде текстовой подписи на экране поверх видеопотока. Дополнительно в интерфейсе выводится информация об ориентации ладони и направлениях пальцев. Экспериментальное тестирование показало, что система работает со средней частотой 24 кадра в секунду при среднем времени обработки одного кадра 42 миллисекунды, что удовлетворяет требованию работы в реальном времени. Общая точность распознавания на тестовой выборке из 86 образцов составила 75,6%. Наиболее высокую точность (~100%) показали жесты «А», «В», «Д», «Е», «И», «К», «Н», «О», также управляющие жесты «ЛАЙК» и «ДИЗЛАЙК». Наибольшие трудности вызвали жесты «Ж», «Л», «П», «Ф», «Ш», «Щ» (точность 50%), что объясняется высокой степенью сходства конфигураций пальцев и недостаточностью используемых признаков.

Разработанный прототип ориентирован на поддержку людей с нарушениями слуха. Используется доступное оборудование – стандартная веб-камера и персональный компьютер, что снижает входной порог для внедрения. Интерфейс максимально упрощён: достаточно запустить приложение и выполнять жесты перед камерой, не требуя сложных настроек.

Дальнейшее развитие системы связано с расширением словаря распознаваемых жестов (до полной дактильной азбуки), добавлением динамических жестов (движений руки), интеграцией с мобильными устройствами (Android, iOS), использованием нейросетевых методов для повышения точности распознавания сложных жестов, поддержкой двух рук и созданием веб-интерфейса на базе TensorFlow.js [2, 3, 8].

Таким образом, применение современных библиотек компьютерного зрения (OpenCV, MediaPipe) и гибридного подхода к классификации (правила + KNN) позволило реализовать прототип системы распознавания жестовой речи в реальном времени. Собранный датасет из 432 образцов для 31 класса жестов, достигнутая точность 75,6% и частота 24 кадра в секунду подтверждают работоспособность предложенного решения. Развитие подобных систем

Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ ЭЛ № ФС 77-68405 ISSN 2541-8327

представляет собой важное направление в области цифровой инклюзии и повышения доступности коммуникации в обществе.

### Библиографический список:

1. Бессонов С. Ю. Распознавание жестов в системах человеко-компьютерного взаимодействия / С. Ю. Бессонов. – М.: Наука, 2020. – 312 с.
2. Гласснер, Э. Глубокое обучение без математики. Том 2. Практика: руководство / Э. Гласснер; пер. с англ. В. А. Яроцкого. – М.: ДМК Пресс, 2020. – 610 с.
3. Ежов А. А. Применение нейронных сетей для распознавания жестов [Электронный ресурс] / А. А. Ежов, В. И. Воронов // Научное сообщество студентов XXI столетия. Технические науки. – М.: МТУСИ, 2018. – Текст: электронный. – URL: <https://scienceforum.ru/2018/article/2018003929> (дата обращения: 15.05.2026).
4. Hand Gesture Recognition using Neural Networks [Электронный ресурс]. – GitHub, 2020.
5. Sign Language Recognition using MediaPipe Hands and Deep Learning [Электронный ресурс]. – GitHub, 2023. – Текст: электронный. – Режим доступа: <https://github.com/nodamu/sign-language-recognition> (дата обращения: 15.05.2026).
6. OpenCV documentation [Электронный ресурс]. – Текст: электронный. – Режим доступа: <https://docs.opencv.org/4.x/> (дата обращения: 15.05.2026).
7. MediaPipe: Solutions for perception pipelines [Электронный ресурс]. – Google, 2024. – Текст: электронный. – Режим доступа: <https://developers.google.com/mediapipe> (дата обращения: 15.05.2026).
8. Scikit-learn: Machine Learning in Python [Электронный ресурс]. – Текст: электронный. – Режим доступа: <https://scikit-learn.org> (дата обращения: 19.05.2026).