

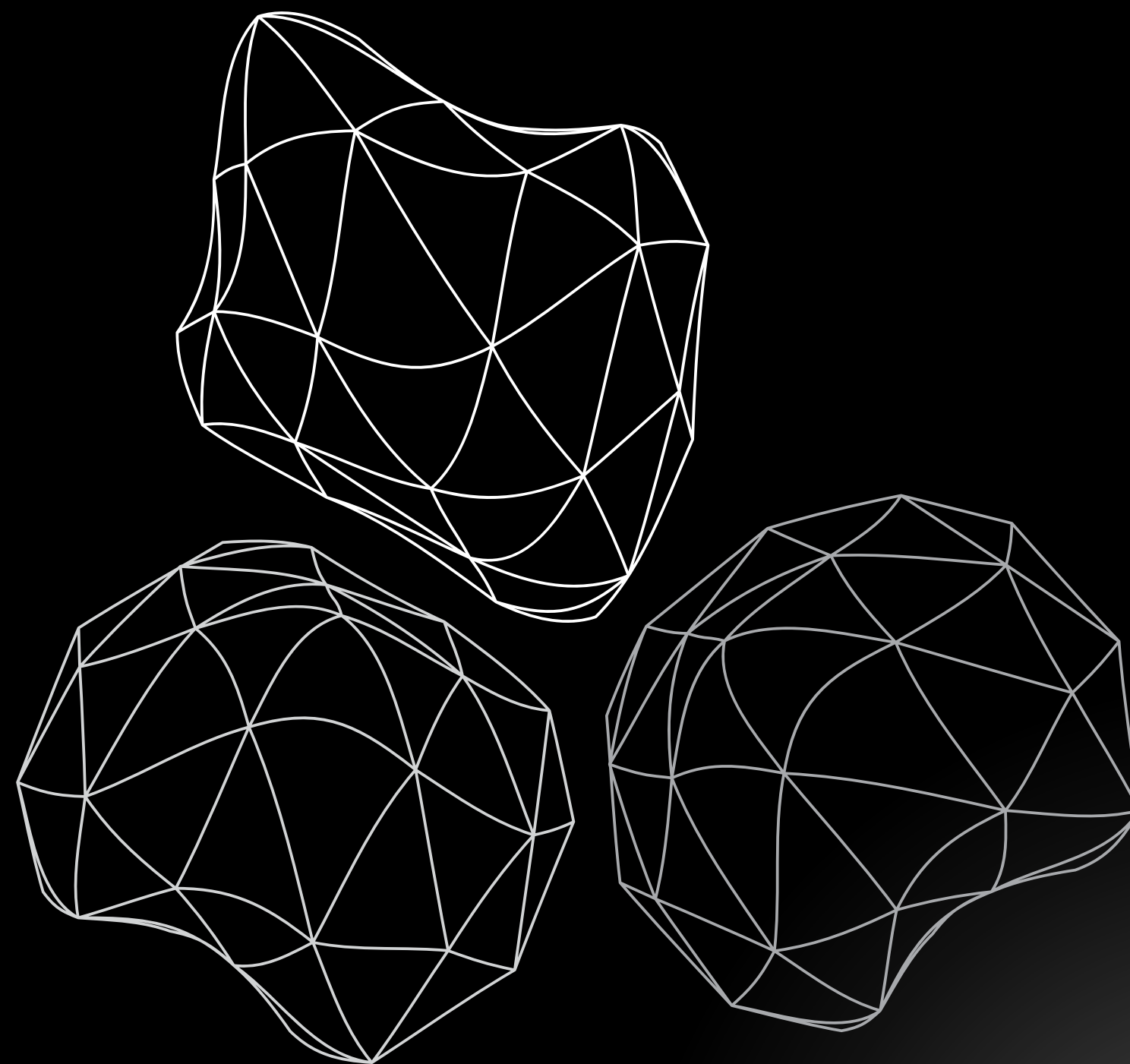


# ОПРЕДЕЛИТЕЛЬ ГОЛОСА

АХМЕТШИН ТАГИР  
ЯКУШЕВА ВЛАДИСЛАВА

# О ПРОЕКТЕ

Проект по распознаванию речи  
и идентификации спикеров  
на основе аудиозаписей.





# СТРУКТУРА ДАННЫХ

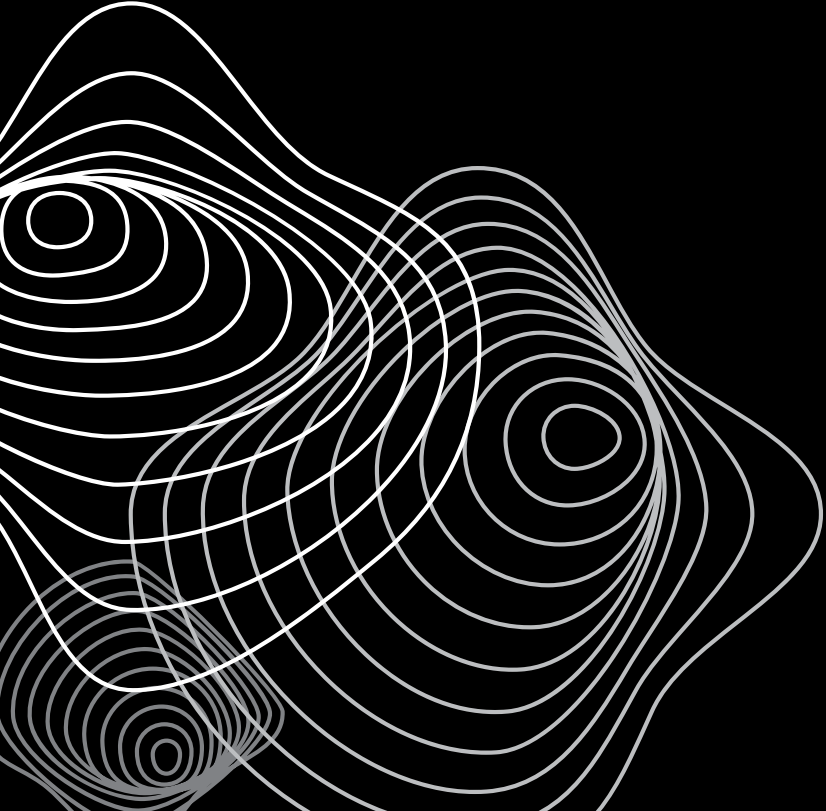
В качестве данных используются аудиофайлы разных спикеров, собранные в отдельные папки.

Данные делятся на обучающую и тестовую выборки. Для повышения устойчивости модели к шуму, в тренировочную выборку добавляются фрагменты фоновых шумов.



# ПРИЗНАКИ

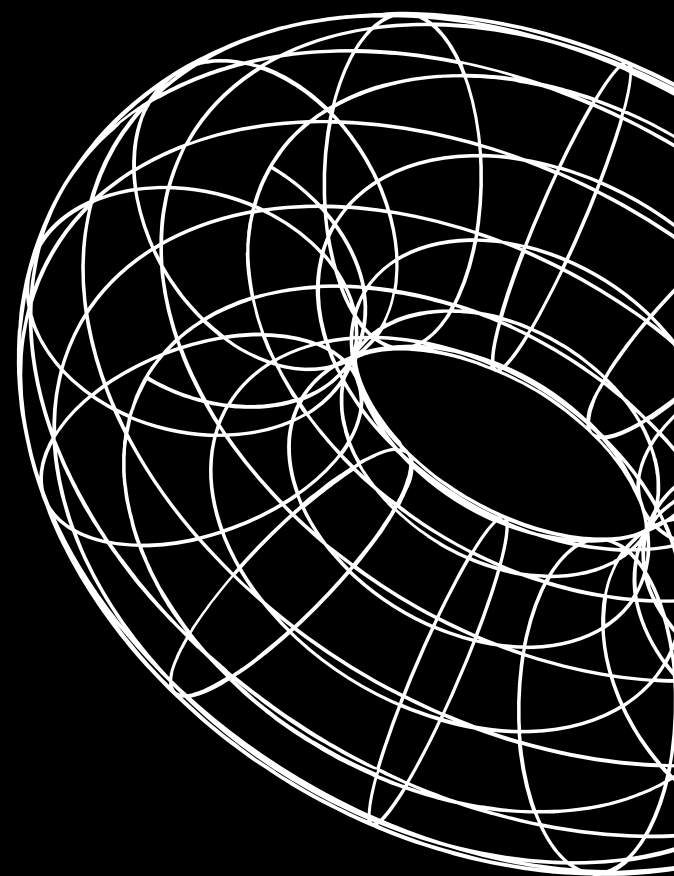
Для каждого аудиофайла извлекается набор признаков: MFCC, дельта-признаки, хрома, спектральный контраст, zero-crossing rate и RMS. Эти признаки позволяют описать структуру и особенности голоса. Такой подход помогает выделить индивидуальные характеристики каждого спикера.





# НОРМАЛИЗАЦИЯ

После извлечения признаков они нормализуются — это важно для корректной работы алгоритмов. Затем применяется метод главных компонент (РСА), чтобы уменьшить размерность данных до 30 признаков. Это ускоряет обучение и помогает избежать переобучения.

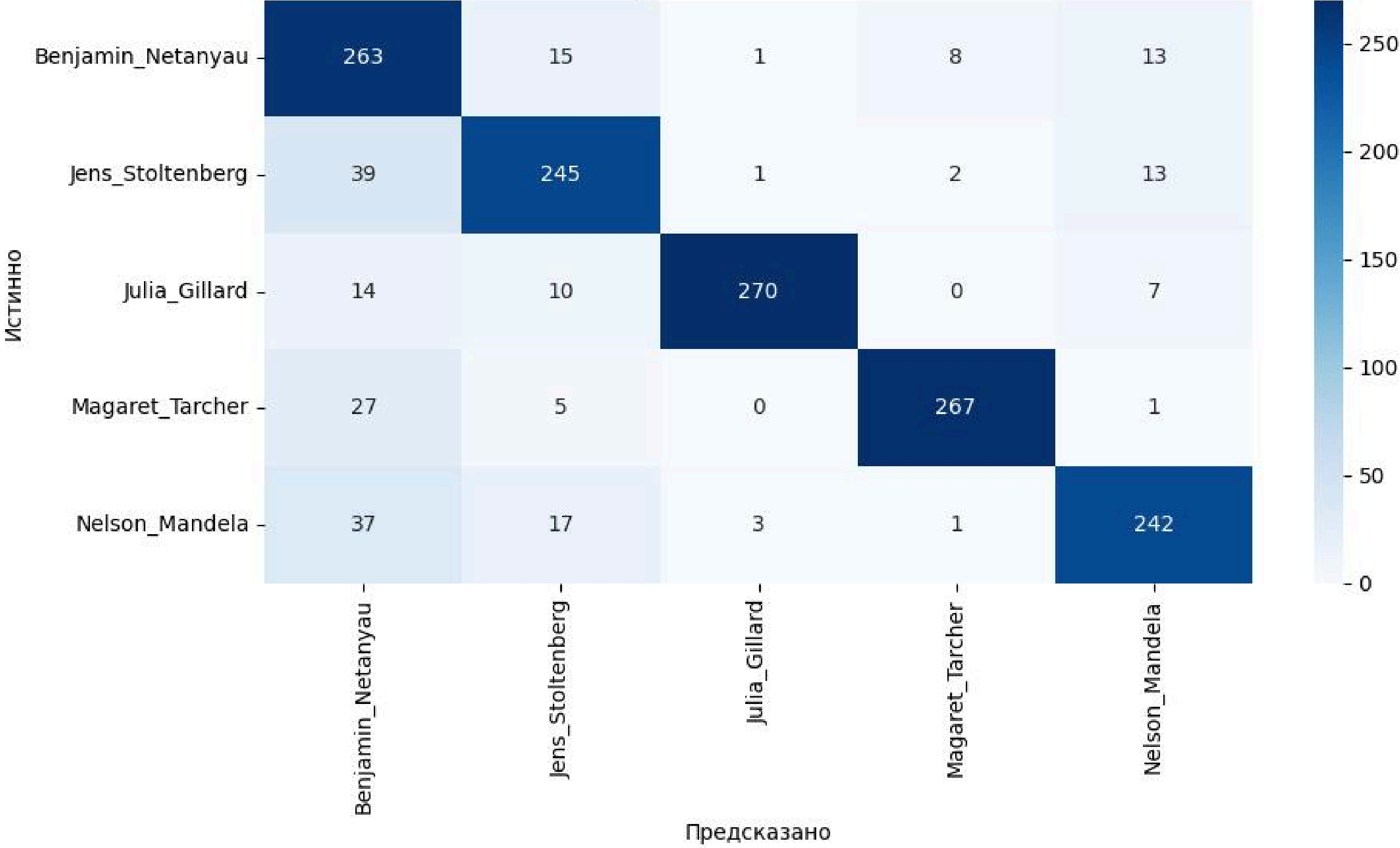




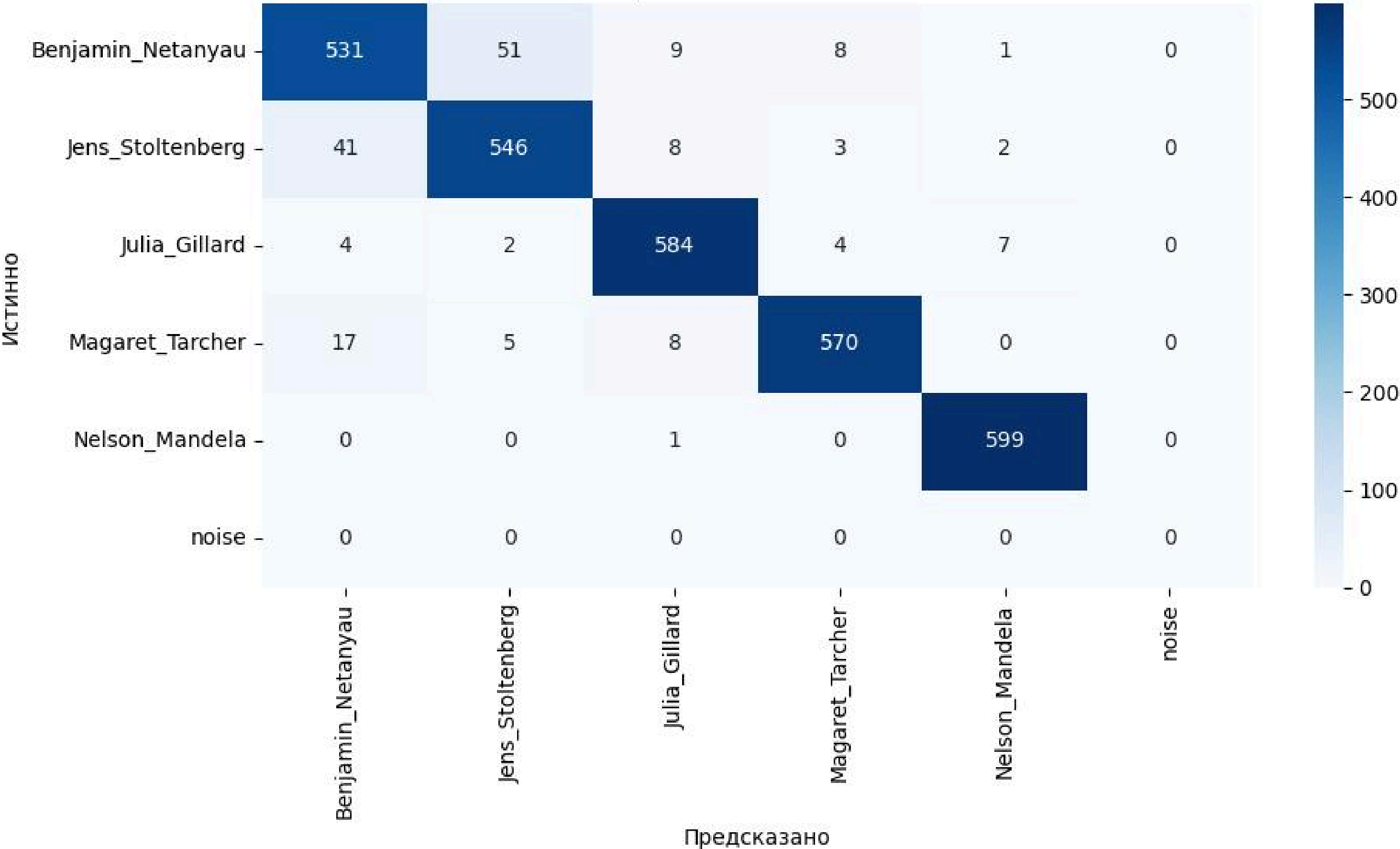
# КЛАССИФИКАЦИЯ

Для распознавания спикера используется алгоритм  $k$  ближайших соседей (kNN). Для каждого тестового примера находятся  $k$  наиболее похожих примеров из обучающей выборки. Итоговый класс определяется по большинству среди соседей.

Матрица ошибок (k = 26)



Матрица ошибок (k = 26)







# ТЕСТИРОВАНИЕ ДАННЫХ

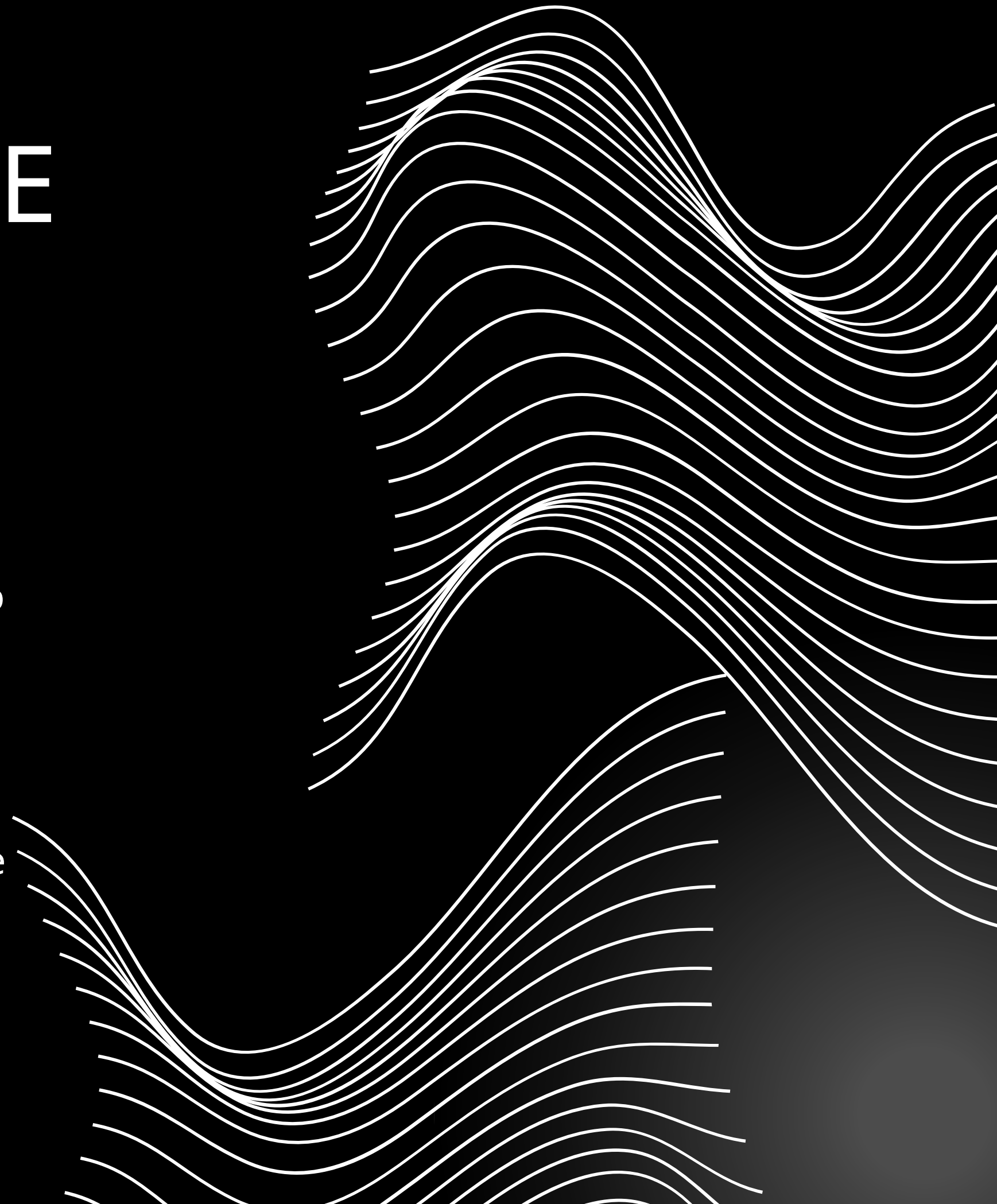
В проекте реализована функция для тестирования модели на новых аудиофайлах. Пользователь может загрузить свой файл и получить результат классификации. Это удобно для проверки работы системы на реальных данных.

# АЛЬТЕРНАТИВНЫЕ ПОДХОДЫ

В проекте реализованы два подхода к идентификации спикеров:

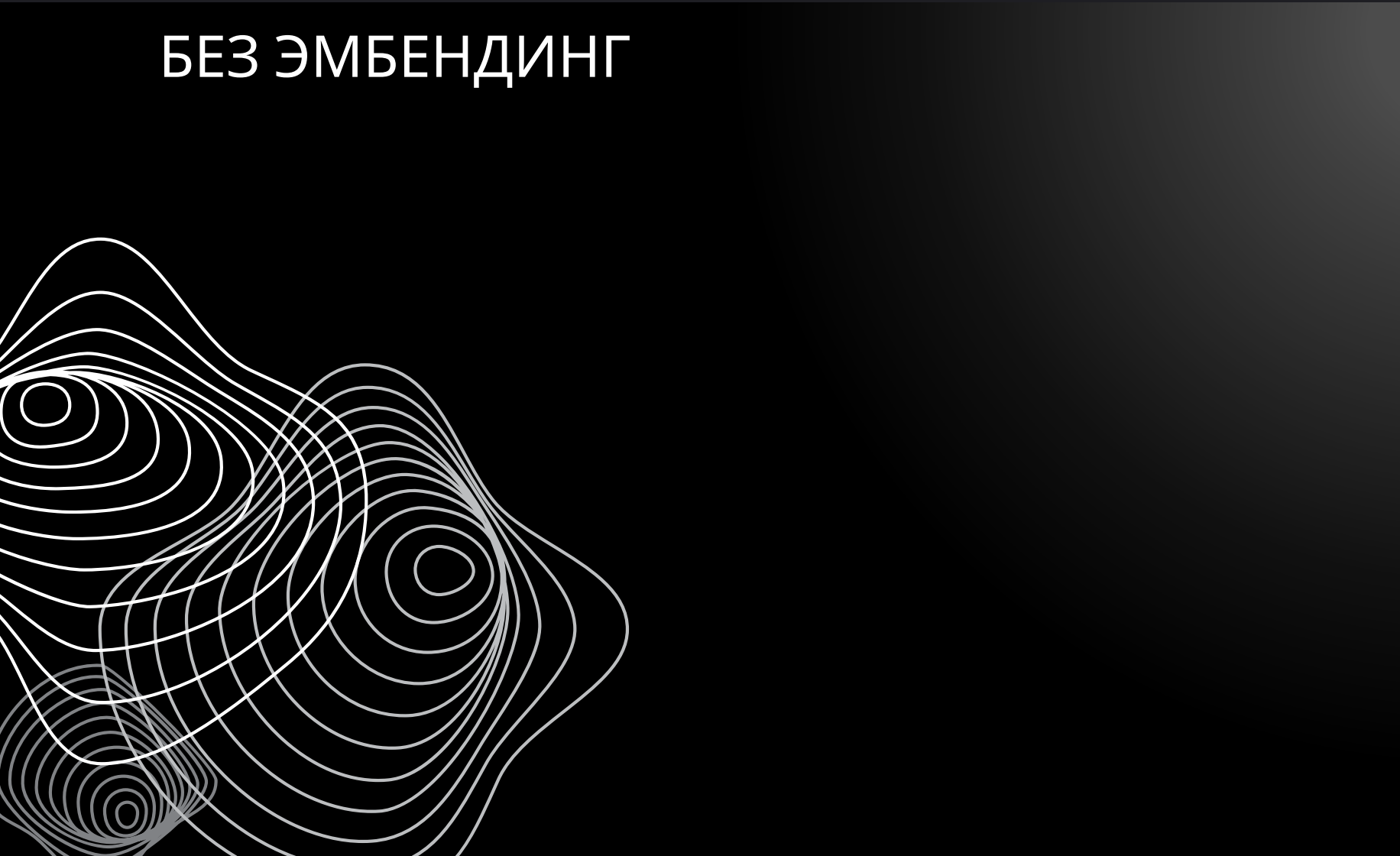
1. Классический: извлечение признаков (MFCC), понижение размерности с помощью PCA или t-SNE, классификация с помощью kNN.
2. Современный: извлечение speaker embeddings с помощью нейросетевой модели (ECAPA-TDNN из SpeechBrain), далее классификация kNN.

Использование t-SNE позволяет визуализировать данные в 2D и выявлять кластеры спикеров.



Точность модели: 0.9540

	precision	recall	f1-score	support
Benjamin_Netanyau	0.91	0.93	0.92	600
Jens_Stoltenberg	0.96	0.91	0.93	600
Julia_Gillard	0.95	0.96	0.96	601
Magaret_Tarcher	0.97	0.97	0.97	600
Nelson_Mandela	0.98	1.00	0.99	600
accuracy			0.95	3001
macro avg	0.95	0.95	0.95	3001
weighted avg	0.95	0.95	0.95	3001

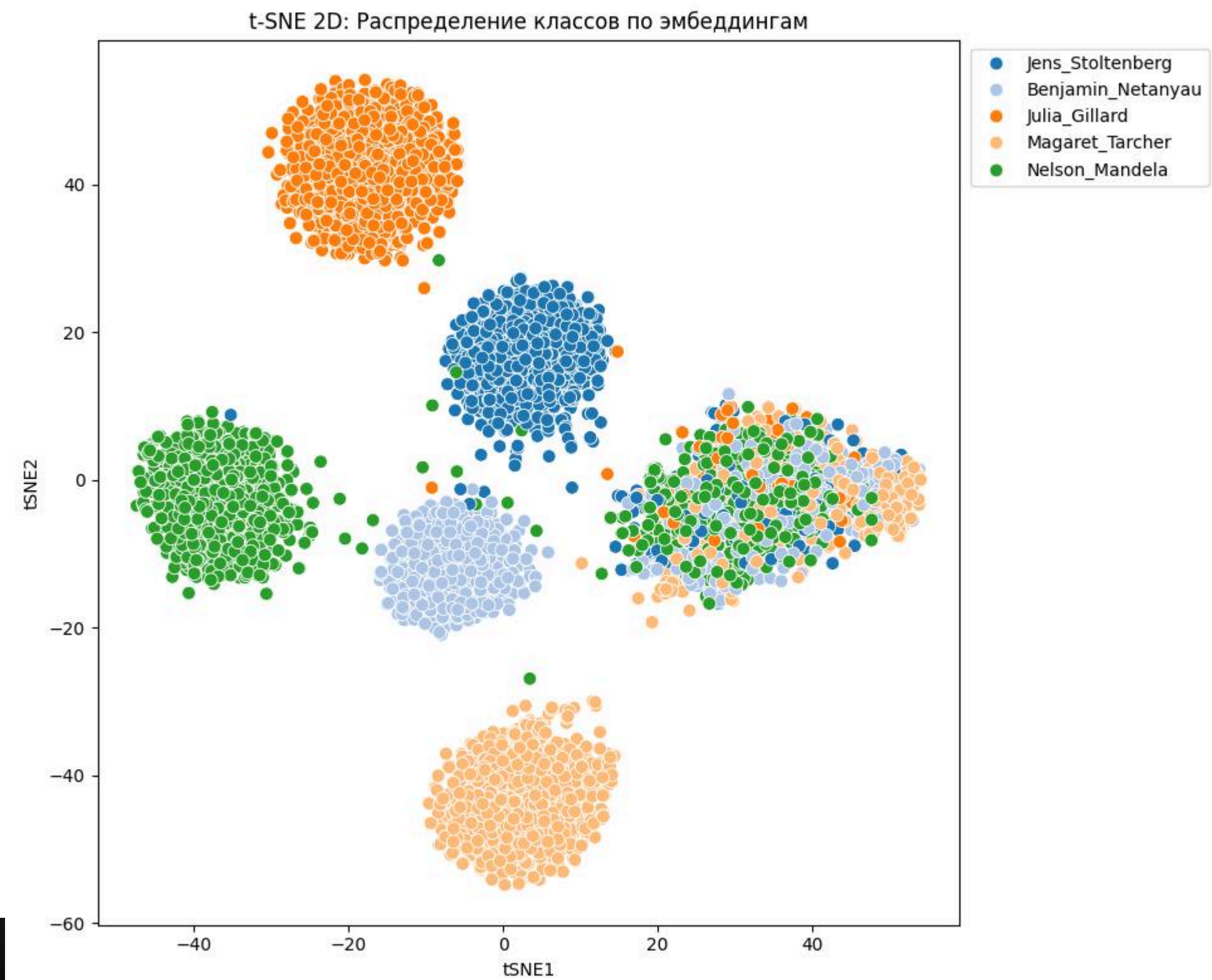
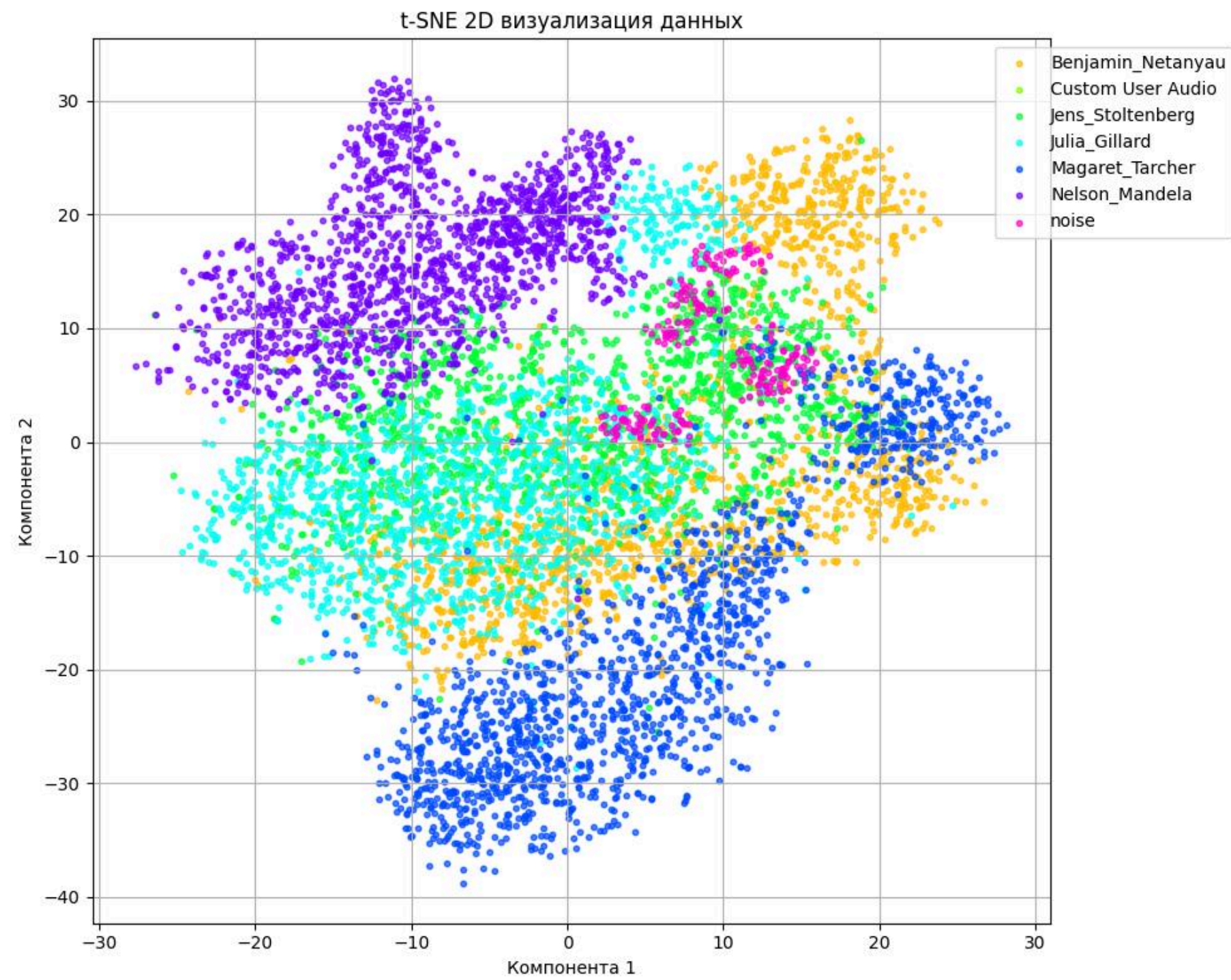


✔ Точность модели на тестовой выборке: 0.8574

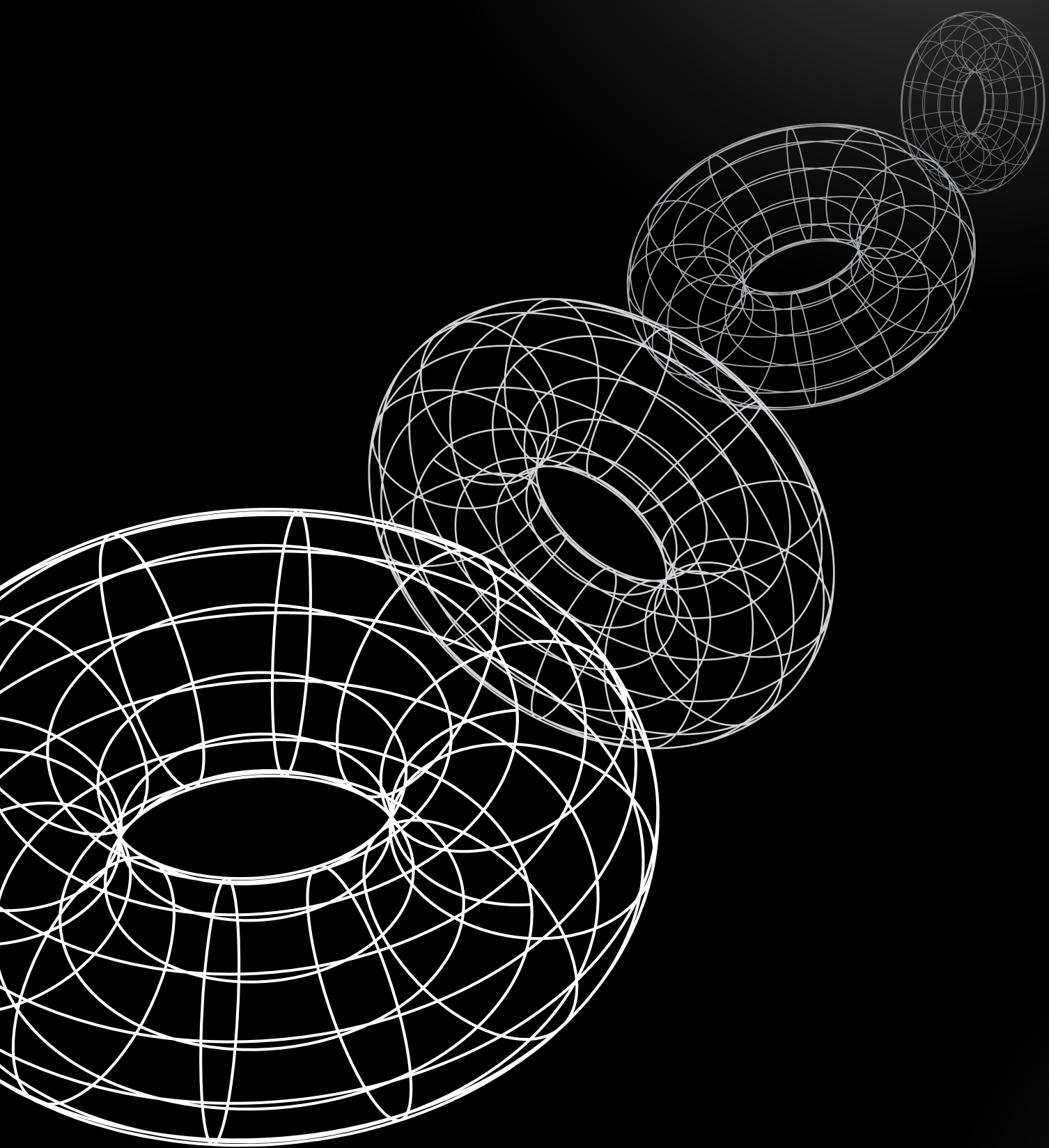
🔍 Классификационный отчет:

	precision	recall	f1-score	support
Benjamin_Netanyau	0.69	0.88	0.77	300
Jens_Stoltenberg	0.84	0.82	0.83	300
Julia_Gillard	0.98	0.90	0.94	301
Magaret_Tarcher	0.96	0.89	0.92	300
Nelson_Mandela	0.88	0.81	0.84	300
accuracy			0.86	1501
macro avg	0.87	0.86	0.86	1501
weighted avg	0.87	0.86	0.86	1501









# ВЫВОДЫ

Разработанная система успешно решает задачу идентификации спикеров по коротким аудиозаписям. В дальнейшем можно улучшить качество за счёт более сложных моделей и увеличения объёма данных. Проект легко расширяем и может быть адаптирован под другие задачи.