

# Speaker recognition

Журавская, Камалбеков, Козловцев  
517 группа

Московский Государственный Университет имени М.В.Ломоносова  
Факультет вычислительной математики и кибернетики  
Кафедра математических методов прогнозирования

24 июня 2019 г.

**Дано:** образцы голосов  $N$  спикеров;  
Обучающая и тестовая выборки;  
У каждого спикера уникальный  $id$ .

**Найти:** для каждого образца в тестовой выборке либо подобрать  $id$  спикера, либо определить, что таких  $id$  в обучающей выборке нет.

- Признаки: mfcc из второго задания!
- Разделим пространство признаков на кластеры: K-Means!
- Найдем порог принадлежности кластеру: если расстояние от объекта до всех центров кластеризации больше порога, то считаем, что спикер новый, т.е. его не было в обучающей выборке.

Детали: Количество признаков 24. Количество кластеров  $4N^*$ .  
Проблемы: плохая разделимость, низкая скорость работы

\*Speaker identification using mel frequency cepstral coefficients

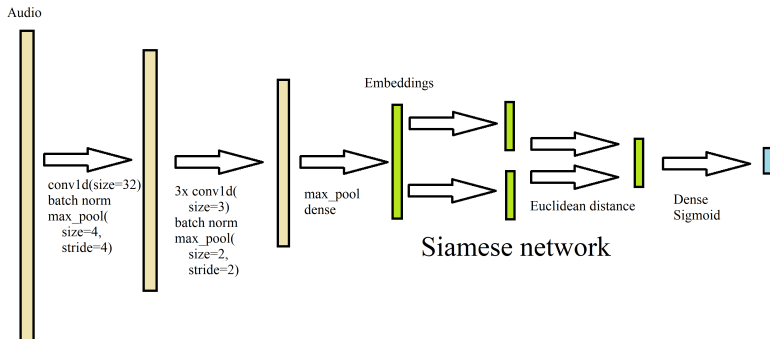
- Данные: VoxCeleb аудио + видео!
- Предобработка: ....
- Сеть: VGGish на GitHub

Проблемы: tf -> torch, свою сеть обучать долго: используем предобученную!

Итог: точность  $\approx 20\%$  на 100 спикерах с наибольшим количеством образцов голоса.

## Building a Speaker Identification System from Scratch with Deep Learning

- Данные: LibriSpeech
- Архитектура:





# Демонстрация?

Мы изучили существующие методы решения задачи распознавания спикера. Мы попробовали несколько методов распознавания на нескольких наборах данных. Получили опыт использования предобученных нейронных сетей.

Наилучший результат получен с использованием siamese networks, точность классификации на наборе данных LibriSpeech ???%!

## Вклад участников:

Саша	идея, наивный алгоритм, презентация
Костя	сеть: выбор архитектуры, перенос весов
Тимур	данные: предобработка и эксперименты