

## Introduction

---

Мы решили сейчас немного отойти от изначально поставленной темы "классификация людей по манере речи" и начать немного с более простой темы. А именно в первые, предположительно, пару месяцев мы будем заниматься задачей VAD (**Voice Activity Detection**). Как понятно из названия, нужно будет по видео фрагменту понимать произносится там речь или нет. Это задача, например, применяется в "Keyword spotting" (в двух словах - это детектирование фраз в духе "OK Google", "Hi Siri"). В классических подходах это делается с помощью аудиофайла и обученной GMM/DNN модели. Мы же предлагаем это сделать с помощью видео и выделенных оттуда признаков с губ.

Теперь собственно с чего начать :)

Самая первая версия VVAD (Visual Voice Activity Detection) будет представлять следующее - вы будете применять уже готовую модель для расставления лэндмарков на лице (ключевые точки, например, точки на губах, глазах итп) и использовать некоторые операции над этими лэндмарками для понимания активности губ (в простейшем случае - расстояния между конкретными точками на губах). Впоследствии эта версия будет использоваться в качестве Baseline для нашей задачи. Для этого

\* Мы предлагаем использовать модель для построения лэндмарков из библиотеки dlib (<http://dlib.net/>). Вам в первое время нужно будет сбилдить ее у себя и попробовать поиграть с исполняемым файлом, расставляющий лэндмарки (webcam\_face\_pose\_ex.cpp; может немного тормозить за счет медленного face detection). Эта библиотека подключена как сабмодуль к текущему репозиторию

\* Как только у вас получится собрать и запустить файл следующим шагом будет построение инфраструктуры для тестирования и обучения моделей. Этот код будет храниться на репо в гитхабе, который я вскоре создам

\* Далее нужно будет найти данные, на которых можно проверить качество вашей простейшей модели (расстояние между лэндмарками). Это могут быть уже размеченные видео файлы, могут быть файлы с субтитрами (по которым можно понимать промежутки с речью), могут быть ряд файлов записанных самостоятельно на свой мобильный телефон

\* В репозитории мы придерживаемся следующего соглашения – master ветка всегда находится в корректном и актуальном состоянии. Как только вы будете планировать влить свои изменения в мастер – вы открываете пуллреквест и дожидаетесь ревью своего коллеги. После этого код попадает в основную ветку