follow-up document for final project:

Automated Lip Reading

**Members:**

Tal Noam - 204397715

Evgeny Vendrov - 313962193

Amitai Zamir - 204324305

**Moderator:**

Lee-Ad Gottlieb

**Table of contents:**

1. Our status at the moment.
2. Our next task.
3. General view of the project.

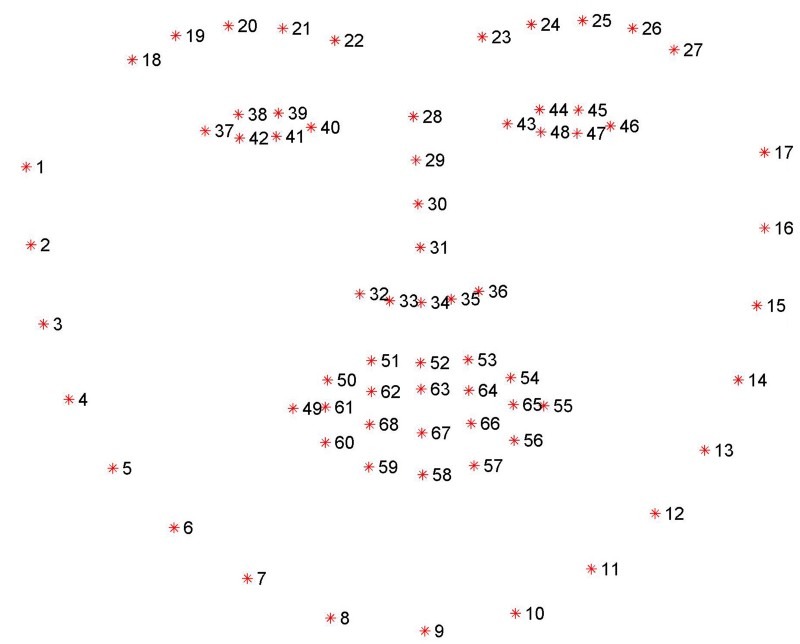
Date:16/03/2020

1. Our status at the moment

Dataset: We have collected a dataset, which size is 15GB.

The dataset contains speaking people (as a "speaking head") – saying basic phrases composed of 3 to 4 words, and, of course, text files which contain everything was said in every video.

Code: We managed to create a python script to frame and highlight the lips of people speaking in these dataset videos, we used a CNN from the "dlib" library to find 68 points in the persons face:



We focus on the 49th to the 68th points, using the library "imutils".

We then, highlight and frame the pixels around the lips using "OpenCV" library.

2. Our next task

Dataset: find videos of people pronouncing only letters, this should be relatively easy as we don’t need to many videos and these videos are widespread over the internet.

Code: we should figure out how frequent – in which rate to capture the "lip frames", and how to "translate" these images to data which the next NN (RNN) could use.

Next, we should start building and training the first version of the complete DNN model which is assembled from a CNN and an RNN, using these "only letters" dataset we have collected.

3. General view of the project

We already finished our work on the CNN which should find the lips and export them as images, out next task will be to learn how to suit that data exported from the CNN to the RNN, later we should find "easy data" (of letter only) to start the training with, and then – finally construct the RNN and start the train phase.