**Final project vision for project instructor Dr.  Lee-Ad Gottlieb**

**Team members:** Evgeny Vendrov 313962193, Amitai Zamir 204324305, Tal Noam 204395715

**Project topic:** Deep Learning Visual Speech Recognition Algorithm

Our goal is to develop an algorithm which will receive voice-less (or will just ignore voice) video of a speaking person as input and will output everything the person said as text, our mission is to tackle this problem as an "open world problem" – we want our system to be able to work well with the whole language, this means – we're not trying to learn only specific phrases or word.

the accuracy of generated text (in comparison to what actually been said) should be high, i.e. the algorithm should be useful and well working.

Our algorithm will work only with English, and as a start we'd try to run it on clear video, without additional noises, as a later desire -we'll try to make it work with not as clear videos i.e. a person which is moving often, a person which is standing in a relatively great distance from the camera, out of sync video, etc…

Our research of Existing products in this field revealed that there are papers which try to solve the "lip read" problem with non-deep learning methods, we found methods using CNN to try and find phonemes or visemes on still images, and there are some deep learning products such as "LipNet" which try to achieve the same goal as we do, additionally, we found this paper [<https://arxiv.org/pdf/1809.02108v2.pdf>] from about a year ago which deals with this problem, so this is possible, and there is already done research to work, and to compare our research to.

As a product, there is a lot of possible uses for this kind of system, consider a lecture, which is filmed (what is really popular), with our tool you could export this video as a text summary, or consider generating subtitle's for streaming services as YouTube (which already has some kind of automatic speech recognition to generate auto captions).