Open Vocabulary Silent Speech Recognition

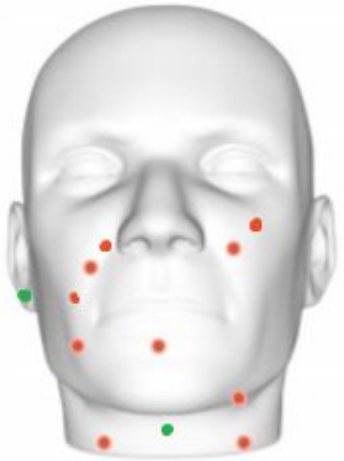
Classical methods of data analysis on raw data and Neural Networks on spectrograms.

Goals and objectives of work:

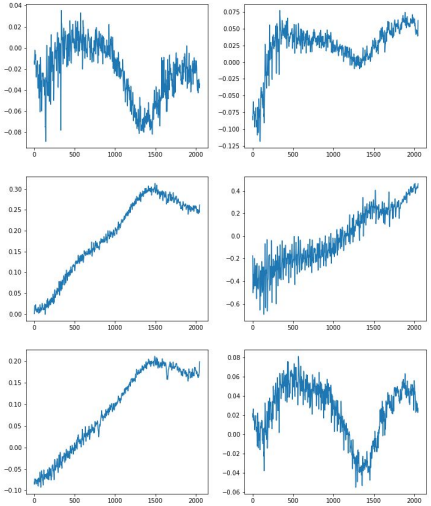
1. Check the conceptual possibility of OVSSR.
2. Collect dataset of phonemes recordings for silent-speech recognition.
3. Spread it in community, so everyone could try their ideas simply.

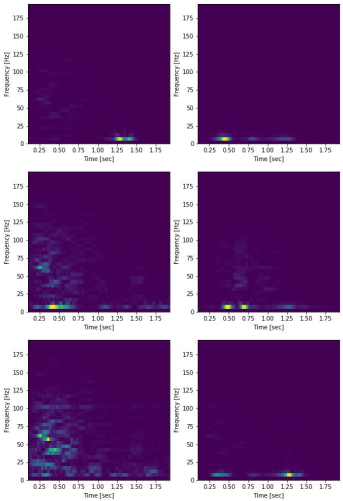
Dataset collection

1. Male english teacher.
2. 30 repeats for every of 44 phonemes of English l
3. 10+2 EMG sensors placed as follows:
4. 1000 hz discretization freq.
5. 500 hz low pass filter.
6. 2050 ms for each sample.
7. No sound or blowing, while recording sample.



Tested methods:





Raw data:

1. Decision tree classifier
2. Random Forest
3. Gradient Boosting
4. KNN-classifier

Spectrograms:

1. Gradient boosting
2. Neural Networks (dense, convolutional, recurrent)

Obtained results:

1. Classical methods on raw data reached 30% accuracy with random forest
2. Gradient boosting - 42%
3. Neural Networks - 29% on dense architecture

Unacceptable overfitting

Random - 2.2%

Conclusion:

Neural networks are hard to tame!
Taking into account results from my colleague (72% accuracy), possible outcomes are:

1. OVSSR is possible on the same conditions as usual speech recognition.
    a. Short-term phoneme recognition
    b. Statistical models applied over results of “a”.

2. OVSSR is possible with personal tuning.

References:

1. AlterEgo: A Personalized Wearable Silent Speech Interface, Kapur et al., April 2018
2. A PHONEME-BASED PRE-TRAINING APPROACH FOR DEEP NEURAL NETWORK WITH APPLICATION TO SPEECH ENHANCEMENT, Shlomo E. Chazan, Sharon Gannot and Jacob Goldberger, 2016
3. Sub-Word Unit based Non-Audible Speech Recognition using Surface Electromyography, Matthias Walliczek, Florian Kraft, Szu-Chen Jou, Tanja Schultz, Alex Waibel, 2006

Thank you for attention!