BRAIN2SPEECH
Interpreting speech from EEG recordings

(BRAIN2SPEECH: Beszéd értelmezése EEG felvételek alapján)

Görcs András
BME
Hungary
andras.gorcs@edu.bme.hu

Józsa Richard
BME
Hungary
jozsarichard@edu.bme.hu

Salamon Ádám
BME
Hungary
sal.adam@gmail.com

Bemutatjuk Az átalunk fejlesztett Brain2Speech agy-gép interfész (BCI)-t amely* koponyán kívüli *elektroenkefalográfiás (EEG) jelek alapján következtet a kísérleti alany beszédére. A feladatot klasszifikációs problémára vezettük vissza. Az megvalósított osztályozó program Convolution Neural Network (CNN)-en és a Kara One adatbázison alapul. A program célja az adatbázisban lévő EEG jelek alapján kapcsolódó kimondani kívánt vagy kimondott szavak megtalálása.

In this paper we are presenting our current state of our Brain2Speech project. This is a scalp electroencephalography (EEG) based brain computer interface (BCI) project. We approach this task a classification problem. In our work is based on the Kara One dataset which is contain EEG and audio recording of the speaker. And we create classifier with Convolutional Neural Network (CNN). Our goal with this neural network to assign the EEG signals to the corresponding word of speaking or thinking. We are presenting the current state of our project. Where we are trying to synthetize words, based on EEG signals.

Keywords—BCI, EEG, Kara One, Speech

I. INTRODUCTION (*HEADING 1*)

In the past decade research groups increased their focus on brain computer interfaces. However, people who need these interfaces for communication can experience slow communication speed. Nowadays researchers achieved really impressive results [1] [2]. These results were mainly achieved by using intracranial EEG which has high signal to noise ratio compared to scalp EEG. However, scalp EEG based systems are more desirable in most cases. [3]

Our Brain2Speech development project is to create a CNN based classifier program which can recognise the word using only the presented EEG signal.

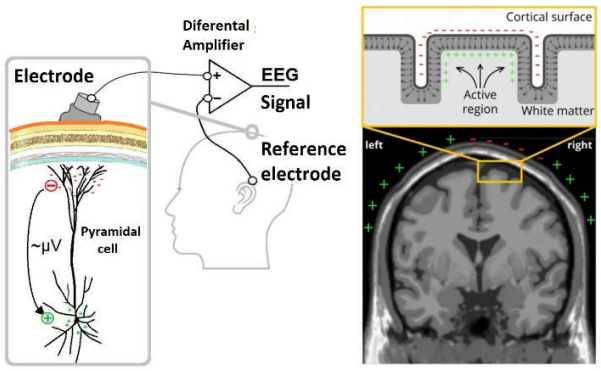
In this paper we are presenting our Brain2Speech BCI program. Our work is based on the Kara One dataset which contains words, extracranial EEG and audio recordings of the speaker. We're creating a classifier using a Convolutional Neural Network (CNN).

II. BACKGROUND

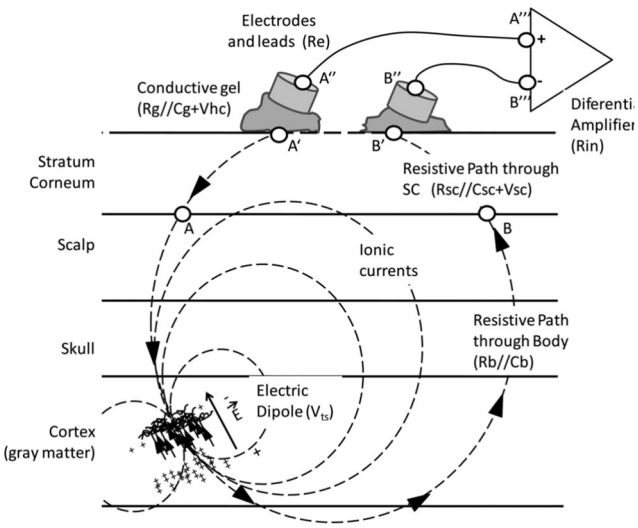
A. Electroencephalography (EEG)

Electroencephalography (EEG) is a commonly used technique to observe the brain activity by measuring the dynamic changes of the electromagnetic field created by the neurons. (1. Figure) Electroencephalography provides a great time domain and weaker spatial resolution of the brain activities. It is able to show the functional state of the brain and its dynamic changes so that it can be studied both globally and in targeted local sub-phenomena of brain activity.

The formation of the EEG signal taken on the surface of the scalp is influenced by several attenuating factors. (2. Figure) Therefore, a synchronous potential change of at least 6-10 cm^2 of cerebral cortex is required to achieve an evaluable signal-to-noise ratio [4] [5], which significantly narrows the range of brain activities that can be studied, but is often used in research and neurological studies.



1. Figure: On the left: a dipole created by the synaptic potential change effect of a stimulating post at a pyramidal cell. On the right: a section of the cerebral cortex that also contains the pyramidal cells. The coming dipoles are summed up during the synchronous activation of the brain area and can be measured by electrodes they create an electric field. [6]



2. Figure: Chematic of electric dipole, ionic currents and differential measure [6].

B. Neurons activity

Neurons show two electrical activities: action potential and post-synaptic potential change. The action potential is triggered when the internal potential of the neuron reaches a value above a threshold level as a result of a stimulus from dendrites. At this point, a self-sustaining potential change of the order of a millisecond extends from the cell to the ends of the axon. The time course and amplitude of the potential change is constant for the given cell. If the activation threshold is reached by the stimulus, it is no longer independent of its parameters. Action potential cannot be detected in most cases with electrodes placed on the scalp. [7] [8] The condition for sensing the action potential change is that several axons run in parallel and that the action potential change in the neurons occurs exactly at the same time. These conditions are present only in auditory-induced cerebral responses. [9] [5]

In our EEG studies, the electromagnetic signals received by the electrodes are typically caused by postsynaptic potential changes in cortical pyramidal cells. [5]

C. Kara One dataset

Kara One dataset, combining EEG and audio records during imagined and vocalized phonemic and single-word prompts. This accesses the language and speech production centres of the brain. [10]

"Each trial consisted of 4 successive states:

1. A 5-second rest state, where the participant was instructed to relax and clear their mind of any thoughts.
2. A stimulus state, where the prompt text would appear on the screen and its associated auditory utterance was played over the computer speakers. This was followed by a 2-second period in which the participant moved their articulators into position to begin pronouncing the prompt.
3. A 5-second imagined speech state, in which the participant imagined speaking the prompt without moving.
4. A speaking state, in which the participant spoke the prompt aloud. The Kinect sensor recorded both the audio and facial features during this stage. prompt.
5. A 5-second imagined speech state, in which the participant imagined speaking the prompt without moving. " [10]

D. Independent Component Analysis (ICA)

ICA [11] solves the blind source separation problem, to recover the independent sources in the signals. Nothing is known about the sources or the mixing process except that there are different sources which produce signal mixtures.

In the case of EEG signals, scalp electrodes pick up EEG signals. With the ICA algorithm the effectively 'independent brain sources' can be generated from the measured EEG electrodes. [12]

ICA can be used in BCI applications to identify and remove noise components by the source. The recorded EEG signals usually contain the noise components originating from blinking, eye movement and muscle activity, their power density spectrum. After removing noises from the decomposed signal, with invert ICA decomposition the remaining useful neural activity can be converted back to create EEG signals reduced by these noises. [13]

III. DATA PROCESSING

The data processing was similar to that described in [14] and was implemented as follows. We use the Kara One dataset [10], which contains EEG signals from 14 participants. There are trials during the measurements. During a trial, the participant first rests for 5 seconds, then a word appears in front of them for 2 seconds, then they thinks about the word for 5 seconds and finally they utter it. Meanwhile, the EEG signals from their head were measured without a pause. From the EEG signal, we exclude VEO, HEO, EMG, ECG, Trigger, M1 and M2 channels, as they are independent or only slightly dependent on speech. Thus, the EEG signal remaining above will have a total of 62 channels. On the remaining channels above, 1 Hz high-pass and 50 Hz low-pass filters are applied. The ICA algorithm is then used to convert the channels into 20 independent components, which are then converted back to the original 64-channel allocation using the ICA inverse operation. Then we segment the EEG signals per trial: we cut out the EEG signal for thinking (thinking EEG) and the EEG signal for speech (speaking EEG). These EEG signals contain different numbers of samples, they have different lengths. However, the neural network expects a fixed length input per EEG type. Therefore, a fixed L value is defined for each type of EEG signal. If the EEG signal is greater than L, we discard the excess samples from the end of the signal, and if it is less, we add samples to the EEG signal until its length reaches L. Based on the lengths of the thinking and speech EEG signals in the dataset, we determined L = 4800 and L = 1200 fixed lengths for the thinking and speaking EEG signals respectively. Finally, the EEG signals are normalized separately. All the tasks described above are performed by the Database class implemented in Database.py. This class is also responsible for downloading and decompressing the data.

IV. Training

Pre-processing yielded thinking and speaking EEG segments (with associated labels: the phonemes thought of or spoken). We also concatenated the thinking and speaking segments to form a new type of training data, as an experiment.

The dataset we used (Kara ONE) consists of EEG recordings of multiple participants. During training participants MM05, MM08, MM09, MM10 were mostly used due to RAM and time limitations (these make up around 25% of all participants), but we also experimented with other combinations.

A. Phase 1 - manual hyperparameter tuning

Our approach can be seen in "train.ipynb".

We started training 1D CNNs separately using these 3 EEG types. Our average net consisted of 3-4 convolutional layers, a few dropout layers between them, followed by a max pooling and a few fully connected layers to top it off. ReLU was almost exclusively used as the activation function, with the last layer using SoftMax, because our approach used classification. We used categorical crossentropy as the loss function and tried Adam and SGD as the optimizer.

We mainly tweaked the convolutional layers' strides, kernel sizes and number of filters, the dropout rates and the fully connected top's size, but we couldn't achieve any significant results.

Thus, we switched to a hyperparameter optimizer tool.

B. Phase 2 - Hyperparameter tuning using Keras Tuner

Our approach can be seen in "hyperparameter tuner.ipynb".

The plan was to implement Keras Tuner's random search, get more familiar with it, build an environment that logs all the necessary data and saves the best models, and then switch to Bayesian optimization, but we didn't manage to do all that due to the lack of time.

We stuck with random search in the end. It crashes whenever it tries parameter combinations resulting in a larger net. We tried handling these crashes the best we could, but couldn't find any sure way to fix them. Anyhow, even this suboptimal way of tuning sped up our work. We logged all the tunings' results, so we could select the best performing models later on.

V. Evaluation

Both Jupyter notebooks contain a section for evaluating the training process and the resulting model as well. Two graphs show how the training and validation loss and accuracy changes over the epochs. This can be used to make further adjustments to the model, e.g. avoid overfitting.

After the training ends, we load the best performing model (based on validation loss) and run predictions on the test data. We compare these predictions to the accurate labels, and calculate the usual classification metrics: accuracy, precision, recall and f1 score. We also visualise the net's predictions using a confusion matrix.

VI. Conclusion, future iprovements

The training should be moved to a workstation with more resources, as this was a huge limiting factor during our work. Using other net architectures could also be useful, for example other researches tried LSTM layers [15], we would like to try them in the future to hopefully achieve better results. Our hyperparameter optimizer environment also lacks lots of functionality, which should be implemented to make the tuning process smoother. The mixing of EEG types, e.g. training on both speaking and thinking segments and testing only on thinking might give us better results due to training on twice as much and more varied data. It could be closer to real world applications as well.

References

[1] C. Pandarinath, P. Nuyujukian és C. H. Blabe, „High performance communication by people with paralysis using an intracortical brain-computer interface," *Medicine, Neuroscience,* 17 February 2017.

[2] D. T. Avansino, F. R. Willett és L. R. Hochberg, „High-performance brain-to-text communication via handwriting," *Nature,* %1. szám593, p. 249–254, 2021.

[3] J. Parvizi and S. Kastner, "Human Intracranial EEG: Promises and Limitations," *Nature Neuroscience,* vol. 4, no. 21, pp. 474-483, 2018.

[4] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq and F. Vidal, "Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view," *International Journal of Psychophysiology,* vol. 3, no. 97, pp. 210-220, 2015.

[5] S. J. Luck, An Introduction to the Event-Related Potential Technique, Massachusetts Institute of Technology, 2014.

[6] B. Farnsworth, "EEG (Electroencephalography): The Complete Pocket Guide," 27. augusztus 2019. [Online]. Available: https://imotions.com/blog/eeg/. [Accessed április 2020.].

[7] M. A. Lopez-Gordo and D. Sanchez-Morillo, "Dry EEG Electrodes," *Sensors,* pp. 12847-12870, 14. július 2014.

[8] Z. Somogyvári, „Az egyedi neuronoktól az EEG hullámokig," [Online]. Available: http://cneuro.rmki.kfki.hu/sites/default/files/NeurontolEEGig.pdf. [Hozzáférés dátuma: 10. szeptember 2020.].

[9] A. Fonyó, Az orvosi élettan tankönyve, Budapest: Medicina könyvkiadó Zrt., 2014.

[10] H. Ábrahám, P. Ács, M. Albu, I. Bajnóczky, I. Balás, A. Benkő, B. Birkás, L. Bors, B. Botz, Á. Csathó, P. Cséplő, V. Csernus, K. Dorn, E. Ezer, J. Farkas, S. Fekete és Á. Feldmann, „Emberi életfolyamatok idegi szabályozása – a neurontól a viselkedésig. Interdiszciplináris tananyag az idegrendszer felépítése, működése és klinikuma témáiban orvostanhallgatók, egészség- és élettudományi képzésben résztvevők számára Magyarországon," Pécsi Tudományegyetem; Dialóg Campus Kiadó-Nordex Kft, 2016. [Online]. Available: https://www.tankonyvtar.hu/hu/tartalom/tamop412A/2011-0094_neurologia_hu/ch09s08.html. [Hozzáférés dátuma: 20. szeptember 2019.].

[11] S. Zhao és F. Rudzicz, „CLASSIFYING PHONOLOGICAL CATEGORIES IN IMAGINED,” August 2014..

[12] A. J. Bell és T. J. Sejnowski, „An Information-Maximization Approach to Blind Separation and Blind Deconvolution,” *Neural Computation,* %1. kötet7, %1. szám6, p. 1129–1159., 1996.

[13] S. Makeig, A. J. Bell, T.-P. Jung és T. J. Sejnowski, „Independent Component Analysis,” *Advances in Neural Information Processing Systems,* %1. kötet8, 1996.

[14] F. S. Rácz, „A nyugalmi agyi konnektivitás multifraktális dinamikája,” 2019.

[15] F. V. a. C. T. G. Arthur, „Deep learning alapú agyi jel feldolgozás és beszédszintézis előkészítő munkálatai.,” in *Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, 2022.

[16] M. J. Monesi, B. Accou, T. Francart és Hugo Van Hamme, „Extracting Different Levels of Speech Information from EEG Using an LSTM-Based Model,” 2021.