<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *10TH IEEE INTERNATIONAL CONFERENCE ON COGNITIVE INFOCOMMUNICATIONS*.

Citation for the original published paper:

Oestreicher, L. (2019)

Cognitive Chunks as Neural Activity: Is it Possible to see What you Think?

In: Prof. Peter Naranyi (ed.), *Proceedings of the 10TH IEEE INTERNATIONAL CONFERENCE ON COGNITIVE INFOCOMMUNICATIONS: CogInfoCom 2019*

Neapel: IEEE conference proceedings

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-396973>

Cognitive Chunks as Neural Activity

Is it Possible to see What you Think?

Lars Oestreicher, Ph. D.
Dept. of Information Science
Uppsala University,
Sweden
Lars.Oestreicher@it.uu.se

Abstract— In this article we present some initial results from the use of Electroencephalograms and Machine Learning to examine the possibilities to find means of communication for people with severe multi-impairments that affect their ability to communicate with other people in their surroundings. The current results from a series of experiments are still tentative, but indicate some interesting possibilities for further development.

In the experiments, informants (without any impairments) have been instructed to look at a screen where simple pictures have been displayed. During the experiment the raw data readings have been recorded and tagged with the corresponding picture that has been displayed at the same time. This data has then been analyzed through a Convolutional Neural Network.

The preliminary result from the experiments is that it is possible to predict which picture the informant is looking at from a single reading of the brain activity with around 80% accuracy. There are further indications of that what we can observe in these patterns is not a representation of the pictures, but rather a representation of a conceptual "chunk". This possibility will be discussed to some end in the Conclusions.

Keywords—Electroencephalogram; Python; Deep Learning Network; Chunk; Picture; Impairment

I. INTRODUCTION

In this paper we will discuss some initial results from a series of experiments set out to study whether it would be possible to support people (primarily children with severe impairments) who lack normal possibilities of communication, through the use of Electroencephalogram (EEG) data in combination with Deep Learning analysis to facilitate a communication channel.

With unskilled eyes the raw signals in an Electroencephalogram (EEG) seem to be almost random. With skilled eyes and special analysis tools, it is possible to detect different kinds of information in the graphs, such as the level of brain activity, stress, focusing, and also medical conditions, such as traces after head concussions and epileptic seizures[1]. But is there additional useful information hidden in the curves winding over the screen? This is the question underlying most of the work described in this paper.

A. Background and purpose

This project originated from a previous project, the MOOMIn-project [2], where children with very severe impairments were given the possibility to play their own music, without any help or support from teachers and personal assistants. The project was very successful, and resulted in many interesting new applications for the children [3], [4]. As one part of the project we also tested a simple EEG-device as possible input to synthesizers for children who had no physical possibility at all to communicate. It turned out that it was much too difficult for the children to manage this kind of simple device *as a music instrument*.

However, at one point we replaced the outlet of this music creation environment. Still, looking at the small game instead, where the aim was to get one of two different circles on the screen to move only by using the brain (this was an application using biofeedback). It turned out that these children were able to switch between moving a green ball and a red ball, using only the will power and a very simple EEG device, in this case the Mindwave Mobile¹. This development spawned a different line of thought which has now developed into the currently reported project, where we look at the possibilities to use simple EEG-devices in order to facilitate communication for this category of impairments.

The starting outset for the currently reported work was to investigate whether it would be possible to use simple, easily recognizable pictures as trigger stimuli for an EEG-based communication, where the person would be training the Deep Learning model on a series of simple, easily recognizable pictures, and then being able to communicate by looking at one of the pictures being displayed in a picture array in front of him or her. Such a system could provide a person with severe impairments with a simple way of telling people about the most important wants and needs, like being tired, hungry, happy or sad, just by looking at the appropriate picture.

With this idea as a starting point, we therefore wanted to investigate the possibility of recognizing patterns in the brain wave occurring when an informant is subjected to a sequence of clearly distinct pictures. For this we used a research grade open source EEG recorder, the openBCI² Cyton board with a Daisy

¹ <http://neurosky.com/biosensors/ee-g-sensor/biosensors/>

² <https://openbci.com/>

attachment, which is able to record raw signal data from 16 separate channels (electrodes) spread over the head at a speed of 125Hz. With this setup we wanted to see if it was possible to get a holistic view of the brain activity rather than using a more localized set of electrodes, or looking at the filtered brain waves, as was the case with the Neurosky MindWave equipment.

B. Hypotheses and short description of the results

The main hypothesis was that we would be able to find some patterns in the data that would at least indicate a connection to the correct picture shown, and also that it would be possible to see the cognitive progression from visual stimuli to the cognitive recognition of the motif as a concept.

There are, so far, two main findings from the experiments in this work:

- It seems to be possible to find patterns in the raw data that relate well to the presented stimuli for the informant. C)
- There was *no indication* of the cognitive progression, (possibly due to delimitations in the hardware/software setup).

For this initial phase of the research we have only had people without impairments (mostly students) as informants. The research setup and the results will be described in more detail later in the paper.

II. THEORETICAL BACKGROUND AND RELATED WORK

General techniques for monitoring brain waves have been developing steadily over a long period of time [5]. However, the combination of Neuroscience and Machine Learning has grown very rapidly only over the last years [6], although we have very limited knowledge about the inner workings of the brain. Within the area of disability research, computer interfaces to neural activity, as well as direct brain-computer interfaces has been developed and used, for example, to control prosthetic limbs [7], [8], and to generate speech directly from the signals in the lower parts of the brain [9], [10]. The phonetic encoding of spoken language also seems to be detectable at some level. In a related research studying the inference of speech on it was possible to see clear distinctions between different phonemes which suggests that articulation does have a special representation in the brain activity [11].

There are a number of similar achievements (see, e.g., [12] for a good overview), which might lure us into the belief that we will have cracked the enigma that our brains constitute within a very close future. However, there is still much more to learn about the basic workings of the brain, and the work reported on in this article has important consequences if it is proven to work, but less stable arrangements for the continued progress.

There are essentially two types of EEG-reading technologies, using either invasive or non-invasive electrodes. The use of invasive electrodes provides more precise control over the signals, but requires electrodes being inserted into the inner workings of the brain, whereas non-invasive electrodes

are instead placed on the skin of the skull, resulting in less discomfort to the subjects, but also some additional experimental problems, such as high susceptibility to noise and contact problems (that some electrodes don't have proper contact with the skull). In this project *only non-invasive electrodes have been used*.

In order to achieve as good a contact as possible, a special kind of gel needs to be used to improve the contact between the artifact and the brain. This fact can of course also affect the situation when the memories are recalled.

The work presented in this article is based on experiments using locally developed Python software, which has been partially described in a M.Sc. thesis [13]. In a second subsection we will describe the Convolutional Neural Network (CNN) that has been developed for the project. (The complete code package used for the work in this paper is available from the author upon request. It is released under the open source paradigm.)

A. Data Gathering Software

The software developed for this project is loosely based on the Python development kit provided by the openBCI Foundation (producer of the hardware). It has been refined and developed to a useful product. The main function of the software is to manage the setup and running of the experiment while gathering the data from the EEG-reader. The main difficulty during the data gathering has been to collect and save the data in synchronization with the display of the pictures without slowing down the looping speed. In addition, software was responsible for the structuring of the data as a preparation for the analysis in the convolutional neural network.

B. The Convolutional Neural Network

The Machine Learning component of the software has been developed in Python, using Keras and Tensorflow for the modelling. The software has been run on an HP workstation with a Nvidia GPU graphics card.

The models used for the CNN are based on a standard model for categorization of pictures in the CIFAR-10 reference data base [14]. The models have then been modified and finetuned in order to get the best results. In order to use this type of network, the data had to be arranged in a certain manner, which will be explained further in the next section. The standard model for the CNN requires that the “images” have the same number of data rows as columns.

III. THE DATA ACQUISITION PROCEDURE

During each trial the informant is placed in front of a special display screen, and the EEG-helmet is placed on the head and fitted to the appropriate size. The connection between the reader and the electrodes are checked rigorously. As soon as the experiment begins the informant is presented with simple pictures selected from a predefined set, on the screen. Each picture is displayed for three seconds each, repeated in random order throughout the whole experiment. For the third set of pictures described below (the alphabet), the session time has

been doubled in order to get a sufficient number of samples for each image instance.

During the presentation of the images, the software records the raw output from the 16 channels of the EEG-device, an *openBCI Cyton/Daisy board*, and tags each data set with the corresponding image (or category, see below) that is being shown to the subject. (Background data is also collected, in case we want to perform some additional analysis later. We have not used this data yet.)

We have used three different types of picture sets during the experiments:

- First picture set: ten simple, easily recognizable pictures of common objects, e.g., an orange, a bicycle, a tree, etc.
- Second picture set: same ten pictures as in the first set, but now mixed up with the corresponding words, making a total of ten categories.
- Third picture set: All letters of the alphabet, in big font sizes, one letter at a time, a total of 28 categories

For each of these categories we have run a number of data gathering sessions of between 15 and 25 minutes, depending on the category size. The channel data has been collected into

Figure 1. Heat maps for three different images made directly from the raw data (normalized). The differences are easy to see visually. As can be seen, the differences between two concepts can be either very small (two left images) or very big (the right image and the two other).

packages consisting of sixteen channels over sixteen time slots, i.e., we group 16 readings into one package. Each package then represents approximately 0,05 seconds, and is tagged with the picture currently being displayed. The intention behind this packaging was to be able to use the data packages as “pseudo images” of size 16 x 16 in order to use convolutional networks for the analysis. This will become clearer when we look at the visual analysis of the data.

For the current results presented here we have collected approximately 30 distinct sets of data, from the same number of informants.

IV. PATTERN ANALYSIS

There has been a large development within the area of machine learning, and not least into the use of different kinds of *neural networks* for the categorization of patterns in large data collections. Since we are looking at large amounts of raw data, the initial attempts of pattern detection have been made using both *Convolutional Neural Networks (CNN)* and *Deeply Connected Networks (DCN)*. The CNN approach turned out to give the best results, so the results presented here are based on the CNN models.

The analysis has been performed in three different ways, 1) an ocular inspection of patterns in the data, where we made heat

maps of the normalized data set, in order to see whether there was any indications of visible patterns in the raw data; 2) a CNN analysis of the data, where the separate time frames for the first 8 was numbered separately, creating distinct categories; and 3) a CNN analysis with only the pictures as the simple categories, without any attempt to separate the perception phases. All of these are discussed in more detail in the next section.

With the original cifar10 dataset, the number of iterations

Figure 2. The accuracy curves for one of the better models. Note that the training accuracy is *lower* than the validation accuracy. The possible reason for this will be discussed further in the text.

(epochs) needed is relatively limited, reaching an accuracy of 90 in around 200 epochs. However, for the “images” in our data sets we have needed to run 4-5000 epochs (without over- or underfitting), in order to reach a result of 80% for the datasets. One explanation for this huge difference is that the differences in the patterns are so small, that they require more calculations to find. Another factor could also be that the CIFAR images are larger “32x32” which makes the process of feature finding slightly easier than when we use “16x16” sized images. Still, as we will see, the results from the CNN analysis is still impressive.

V. RESULTS

The results after the first rounds of analysis were surprisingly strong. In fact, in order to confirm that there were no methodological errors, a “debunking seminar” was held, where a number of colleagues, not participating in the project, were invited to tear the results apart. However, no methodological error was found.

A. Visual analysis of raw data

In order to see what kind of analysis to perform we started by producing a simple heat map representation of the normalized raw data presented in square shapes. These heat maps are in most cases relatively easy to distinguish from each other by just looking at them (see figure 1). The colour set chosen for the heat maps is chosen at random.

As can be seen in the figure, the most significant pattern is across the rows, but it is possible to see significant fluctuations over the columns. The different images for the same picture are constant, regardless when the picture is displayed during a session. This, at least indicates that there is some regularity to the patterns in the EEG for different images worth investigating

with a CNN. In the next two sections the results from the analysis with the neural networks will be described.

B. Traceable nerve signal progression

When it comes to the analysis of the data in the CNN, the result mainly shows that with our equipment, there is no possibility to see any distribution of the signal over the visual cortex. Regardless of any changes to the recording and analysis models, there was no data set found that proved to have this kind of significant development over time for the same picture. Instead the patterns did show some smaller fluctuations, persistent over time (see figure 1). To catch these fluctuations, however, will be important for the general applicability of the analysis (see next section).

C. Images as patterns

The results from using the deep learning network to try to recognize the images that the person has looked at when the signals were recorded were more interesting. The basic hypothesis was that if there would be any patterns occurring at all, they would be of low accuracy and also difficult to differentiate from each other. This hypothesis was revoked fairly early on.

1) Session runs

In figure 2 one of the accuracy curves is shown for one data set and one of the more successful models. In this run, we reached an accuracy of around 80%. As can also be seen is that the curve is not smooth but rather jagged. However, it can also be seen that the training accuracy is *lower* than the validation accuracy. This is slightly surprising, since it would be expected to be the other way round. There are, however, no indications of over- or underfitting in the data. A possible explanation of this unexpected result is that the CNN uses high levels of dropout in order to avoid overfitting of the curves. Thus, the final result shows that the model is finding the features of the respective brain activity patterns.

It is clear from both the accuracy curves and from the confusion matrix, that it is possible to detect which picture a person is looking at by analyzing the brain activity during that time. We can also see that when there is a mistake in the classification, the confused pictures have some common feature. This becomes even more evident in the next section, where we look at the possibility to recognize not only pictures, but the concepts themselves.

2) Delimitation

So far, we can analyze one data set, and predict the patterns within that data set, all data collected at the same session. One plausible explanation for this may be that the EEG-equipment used is not that precise in its physical placement on the head as needed to find the same pattern again. Unfortunately, this has the effect of decreasing the practical applicability. However, we might still think of using a continuous stream, first used to calibrate the model, and after that, without changing the settings, as a communication tool.

Currently we are looking at using a different type of cap for the electrodes that allows for more precise placement of the electrodes on the head from session to session. Hopefully this

will make the models created reusable for the same individual over time.

D. Pictures or concepts?

One extension of the experiments was made by mixing the pictures shown with the names of the object during the presentation, still placing both in the same category. We now have 20 different displays, but still only ten categories. The reason for this variation was to see whether it really was the actual picture that gave rise to the pattern we have detected.

When we ran the new series of experiments, we reached almost (only slightly lower) accuracy as with just the pictures. This was a bit unexpected, and we will discuss this result further in the next section.

VI. CONCLUDING DISCUSSION

The results of this study are surprising in the way that we can use very simple methods and get this kind of accuracy from the analysis of the brain activity. We have cross-checked both the data gathering process and the analysis process in different ways, and have so far not been able to find any methodological errors in the experiments.

There are a number of possible conclusions that can be drawn from this initial study, and we also hope that we will be able to take these studies further in the near future.

A. Brain Patterns

The first immediate conclusion from these initial studies is that it seems to be possible to detect the effects of the presentation of different visual stimuli to a person in terms of patterns that reflect the shown picture. 8 times out of 10 the CNN was able to connect the pattern to the corresponding picture. Although this is not a perfect score, it still means that we are well over a random distribution, and we are currently working on how to increase the accuracy further.

Still, even if the accuracy could be improved, the current accuracy is sufficient for the construction of the communication tool for the children that we are targeting in the project. The real time frame rate is still within 10 samples per second, and with 7-8 correct predictions out of 10, it means that one second of real time recordings would be sufficient to assess the picture that the person is looking at, by a simple "majority vote" over the samples taken during a second.

So, assuming that we can increase the precision of the physical placement of the electrodes, there already seems to be a possibility to develop this type of EEG-recording device into a simple communication device as sketched initially.

So far, we have not been able to transfer the trained models from the data for one person to the data from another person. This means that we cannot train a model on one person and use it to match patterns from another to the pictures they are seeing. This might be seen as a problem, but it can also be positive, as will be discussed in the last section. As of now, we can only create models on a personal basis. Still, this is not a problem for the intended usage of the results.

B. Conceptual chunks?

As mentioned in the previous section we did a small series of studies where we used both pictures and the corresponding words in text as stimuli. The accuracy turned out to be almost the same as for the plain pictures, which was quite surprising. However, this indicates that the patterns are not directly connected to the pictures themselves, but rather to what is evoked in the brain by the visual stimuli. It is still, too early to draw any wider conclusions from the small set of experiments that have been made with this setup. However, if we consider the initial idea with this project, namely to use a holistic perspective on the recording of the data, we might at least suggest some explanations that we intend to investigate further.

Within cognitive psychology, a knowledge object is often referred to as a “chunk”. A chunk can be any meaningful unit that can be referred to. This means that a chunk can be just a single letter, or the whole universe [15]. This is a very convenient model for the storage of knowledge in the brain, since it supports the idea of an associative learning process and storage. A chunk is in some sense defined by the knowledge it is linked to during the learning process.

This also links in well with the observation that the weighted models are not transferrable from one person to another (although the analysis model in the CNN will be the same for all sessions). If we assume, for example, that a chunk that represents a strawberry is constructed from a long learning process, a process that is different for everybody (since we all have different background knowledge to link the new knowledge to) it makes sense that the patterns evoked in the brain for a certain concept will not be the same for different persons. Thus, if we are truly “seeing” the concepts, it would not be any surprise if the analysis for one person would not be transferrable to another person’s concepts. Still, this hypothesis needs a much more detailed investigation in order to confirm or reject.

VII. FUTURE WORK

The results achieved so far are intriguing and invites to a number of follow-up studies. We are currently in the process of launching a larger project that will aim at two major developments:

- We will try to improve the accuracy and the precision of the physical equipment, so that the results will be more stable over time for the same individual. Currently, there are difficulties of transferring models between different periods of time. This increased precision may also make it possible to transfer weighted models between persons.
- We will extend the experiments to investigate the hypothesis of a conceptual representation in the patterns. In this work we would add other dimensions of stimuli, such as spoken words, to see whether the results are changing or not.

In parallel, we will continue the more applied process of developing the communication tool that we hope to be a final result of this project.

VIII. ETHICAL CONSIDERATIONS

As a last comment we want to stress the ethical aspects that constitute a large issue within this project. The brain is probably the most personal of the organs in the body, and trying to access information embedded in a person’s mind naturally raises deeply ethical questions. This becomes even more difficult with the current target user group, since they do not have the communication possibilities as everybody else.

One immediate issue is that of breaching the privacy of the subjects, which becomes an even more important issue when we consider the final target group for the application we aim at developing.

There is also the aspect of “Big Brother sees you”. If we can see the concepts that a person is thinking of, then it would be possible to read the minds of anyone, using this method. This could lead to unforeseeable consequences, such as when scanning for bomb threats, there will be a good reaction in this little valley.

However, as stated in the results section, so far, we are only able to analyze and predict the images from the same person’s data. This is, actually, good for both considerations above, since the person has to give consent for a recording before the analysis can take place, and likewise, it will not be possible to apply a model taken from one person on the patterns from another. This means that both these ethical problems will be diminished quite dramatically.

With the intended target group, there is of course still the problem of getting consent from a person who cannot communicate, but we hope that this will be possible to solve when we have developed the project that far. In the best of cases we would even be able to ask the participants themselves for consent during the first sessions using the equipment as we wish it to be used.

IX. ACKNOWLEDGEMENTS

I am very thankful for my heads of unit, previous and current, for supporting my work, both practically with hardware and scientifically with constructive comments and suggestions. I also want to thank my colleagues at the unit for the support, both with the participation in my “debunking seminar”, and help with discussions and practical help with methods and software.

I also want to thank the students who have written some of the software, which has allowed me to focus on the more theoretical issues of the theses. Furthermore, without informants it is impossible to perform experiments.

Finally, I want to thank the children at Årsta special training school, who have inspired me to continue this work, just by being the wonderful children they are.

REFERENCES

- [1] S. Noachtar and J. Rémi, “The role of EEG in epilepsy: A critical review,” *Epilepsy and Behavior*. 2009.
- [2] L. Oestreicher, “The use of Information Technology for Creative Activities in Special Schools for Children with Severe Impairments - the MUMIN project.” (submitted manuscript)

- [3] A. Rubensson, “*Visualization of Music Designed for Children with Severe Disabilities*,” M.Sc. Thesis, Uppsala University, 2017.
- [4] I. M. Satrya Rudana, “*Kinect-based Music Application for Children with Severe Physical Disability*,” M.Sc. Thesis, Uppsala University, 2016.
- [5] M. Teplan, “FUNDAMENTALS OF EEG MEASUREMENT,” *Meas. Sci. Rev.*, vol. 2, no. 2, 2002.
- [6] A. Riccio, D. Mattia, L. Simone, M. Olivetti, and F. Cincotti, “Eye-gaze independent EEG-based brain-computer interfaces for communication,” *J. Neural Eng.*, vol. 9, no. 4, p. 045001, Aug. 2012.
- [7] R. A. Andersen, S. Kellis, C. Klaes, and T. Aflalo, “Toward More Versatile and Intuitive Cortical Brain-Machine Interfaces,” *Curr. Biol.*, vol. 24, no. 18, pp. R885–R897, Sep. 2014.
- [8] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran, “Brain-Controlled Interfaces: Movement Restoration with Neural Prosthetics,” *Neuron*, vol. 52, no. 1, pp. 205–220, Oct. 2006.
- [9] K. Mohanchandra, S. Saha, and G. M. Lingaraju, “EEG based brain computer interface for speech communication: Principles and applications,” *Intell. Syst. Ref. Libr.*, vol. 74, pp. 273–293, 2015.
- [10] H. Akbari, B. Khalighinejad, J. L. Herrero, A. D. Mehta, and N. Mesgarani, “Towards reconstructing intelligible speech from the human auditory cortex,” *Sci. Rep.*, vol. 9, no. 1, p. 874, Dec. 2019.
- [11] A. Kovacs, I. Winkler, and K. Vicsi, “EEG correlates of speech: Examination of event related potentials elicited by phoneme classes,” in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2017, pp. 000115–000120.
- [12] L. Izso, “The significance of cognitive infocommunications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with non-standard cognitive characteristics,” in *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2015, pp. 77–82.
- [13] T. Menchini, “Pattern detection in Brain Signals from Simple Visual Stimuli,” Uppsala University, 2019.
- [14] A. Krizhevsky, “CIFAR-10 and CIFAR-100 datasets.” [Online]. Available: <https://www.cs.toronto.edu/~kriz/cifar.html>. [Accessed: 15-Jun-2019].
- [15] J. R. Anderson, *Cognitive Psychology and its Implications*, Seventh Ed. Worth Publishers, 2009.