EEG Semantic Decoding using Deep Neural Networks

Juan Manuel Mayor Torres, Evgeny A. Stepanov & Giuseppe Riccardi

University of Trento - Department of Information Engineering and Computer Science

juan.mayortorres,e.stapanov,giuseppe.riccardi@unitn.it

Introduction

The neural representation of semantic categories is still considered a key question in cognitive neuroscience. Differences between conceptual representations are detectable in **early neural activations** [2], suggesting a specific functional organization of concepts that is reflected in neural responses [4] (e.g. Tools and Mammals). There is still an incomplete understanding of the early activations associated with meaningful stimuli, whether pictorial, orthographical or auditory. This limitation is associated with the low number of signal features in each trial, the limited number of training examples, and the high level of inter-subject variability.

A reduced inter-class separation (between **semantic neural responses**) is more evident when the classification is evaluated independently from the subject or involves **single unseen exemplars** (e.g. leave-one-trial-out) [7]. On this specific point, some studies have focused on decoding concept-related brain activity based on features from the stimulus onset (typically from 0-1200 ms) or a denoised long EEG trial where the multiple features obtained have non-deterministic distributions.

Main Objectives

1. Improve the semantic class recognition performance, giving early neural responses as a training-set
2. Recreate realistic scenarios such as Cross-Subject:***Leave-One-Subject-Out*** (LOSO), Within-Subject: ***5-Fold per subject*** (5-Fold), and ***Leave-One-Trial-Out per subject*** (LOTO).
3. Use minimum amount of training examples for each evaluation and modality in order to consider the system independent from extra information.
4. Extend typical ***Deep Neural Networks*** (DNN) systems in order to learn fro different features distributions as Bhattacharyya Distances and PCA [3], showing a considerable performance with a statistical significance between results in posteriors ROIs.

We propose a powerful combination of Bhattacharyya distance criterion mapping and a Deep Neural Network (DNN) classifier to improve the performance of concept decoding for within-subject analysis, and a cross-subject analysis. A complete pipeline of this evaluation is shown in detail in Figure 1.

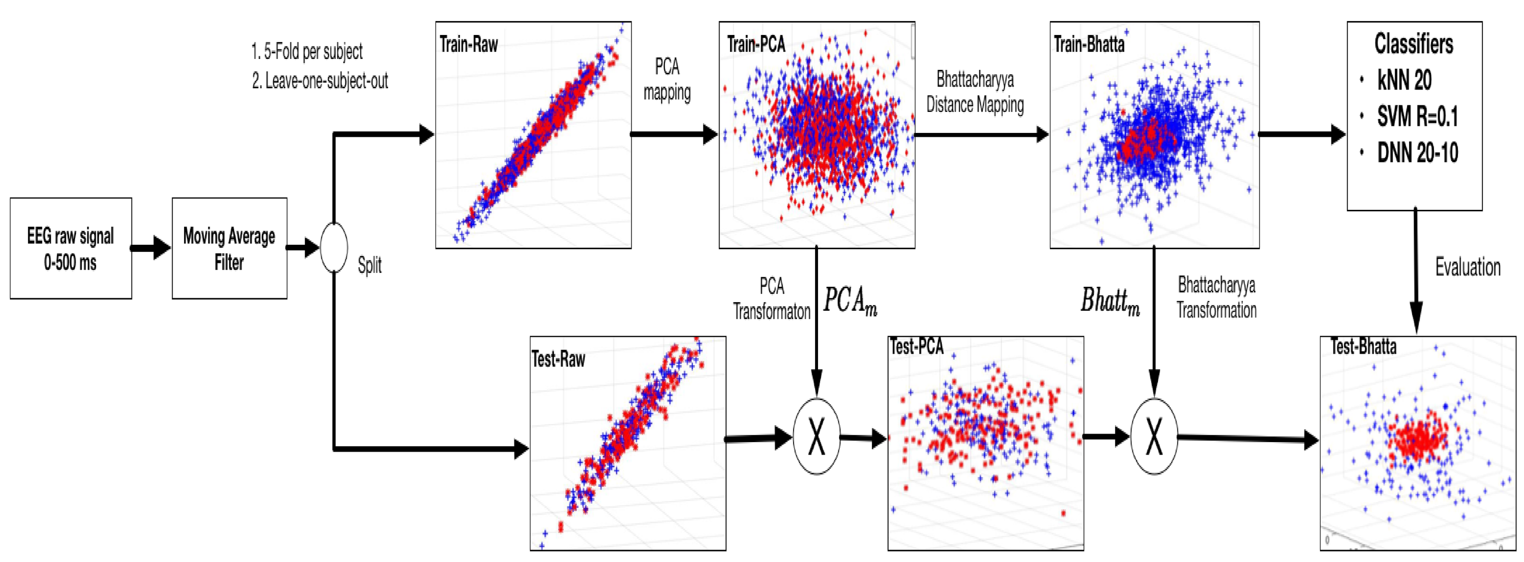

Figure 1: Semantic decoding pipeline and its corresponding subcomponents. The training and test distribution evolves through each new feature space (PCA and Bhattacharyya), obtaining concentric separation regions at the end of the entire process

Materials and Methods

Dataset and Experimental Setting

The data had been recorded from seven healthy Italian speakers (5 male and 2 female,mean age $\mu = 29$), who were asked to silently name animal and tool objects presented in normalized grey-scale photographs. There were 30 land-mammals and 30 work-tools photographs, each presented times in random order for each participant, consolidating a total of 180 trials per class, 180 for Tools and 180 Mammals respectively.

More detailed aspects of these recordings are defined in [6] and its supplementary material.

We divided our seven critical Regions-Of-Interest (ROIs) following [5]. Figure 2 shows this distribution in detail.

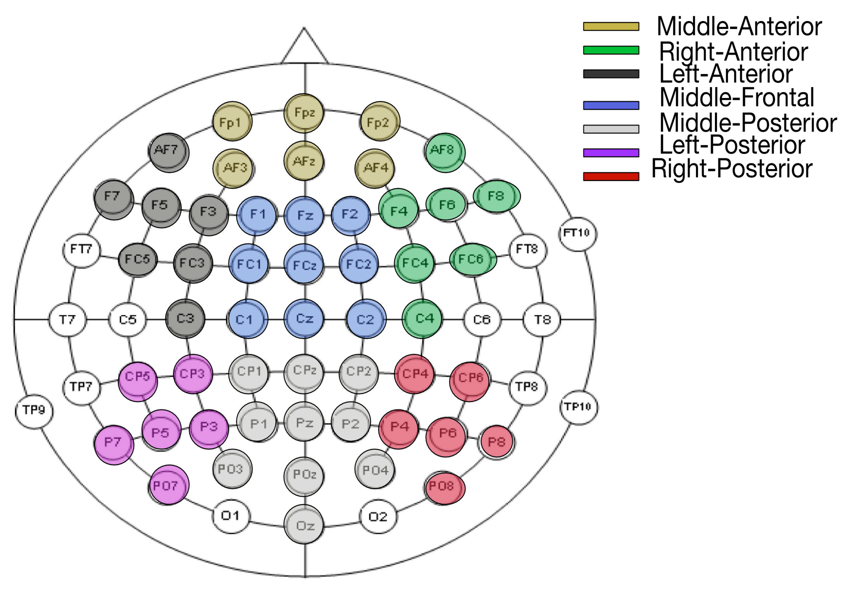

Figure 2: Channels distribution per ROI

Analysis

This methodology is composed of 3 main stages. First a pre-processing stage that consists in a moving average filter application for all the 64 channels in the trial. Second, we split the data compound in : (a) a leave-one-subject-out (6-to-1 cross-validation), (b) 5-fold cross-validation per each subject (4-to-1 cross-validation) following the steps proposed in [6] to generalize unseen exemplars for each semantic category , and (c) an additional LOTO per subject modality in order to extend a generalized model of the PCA+Bhattacharyya+DNN pipeline.

DNN training

1. Using a PCA and Bhattacharyya distance criterion based on the *sequential search*, we calculate the best 20 ranked features for a new feature-set mapping to feed kNN, SVM and DNN (2-layer: 20 input units - 10 hidden units and 2 binomial units for output) classifiers in our evaluations.
2. Each DNN unit is a ***Restricted Boltzmann Machine*** (RBM). Thus, we use a pre-training based on a auto-encoder scheme and *unsupervised Contrastive Divergence* of 100 iterations [1], using a $\epsilon_{CD} = 0.1$.
3. Consecutively, we execute 2500 iterations of fine-tunning back propagation to re-calculate the weights inside the neural network. We use a $\epsilon_{fine} = 0.01$

Results

In these approach we replicate the modalities described by [6], using a *cross-subject* modality or a Leave-One-Subject-Out (LOSO), *within-subject* modality 5-Fold per subject and an extra modality such as Leave-One-Trial-Out (LOTO) per subject that is not included in this evaluation. Refer to Table 1 for the baseline and our results based on kNN-20, SVM R=0.1 and DNN 20-10 two layer classifier.

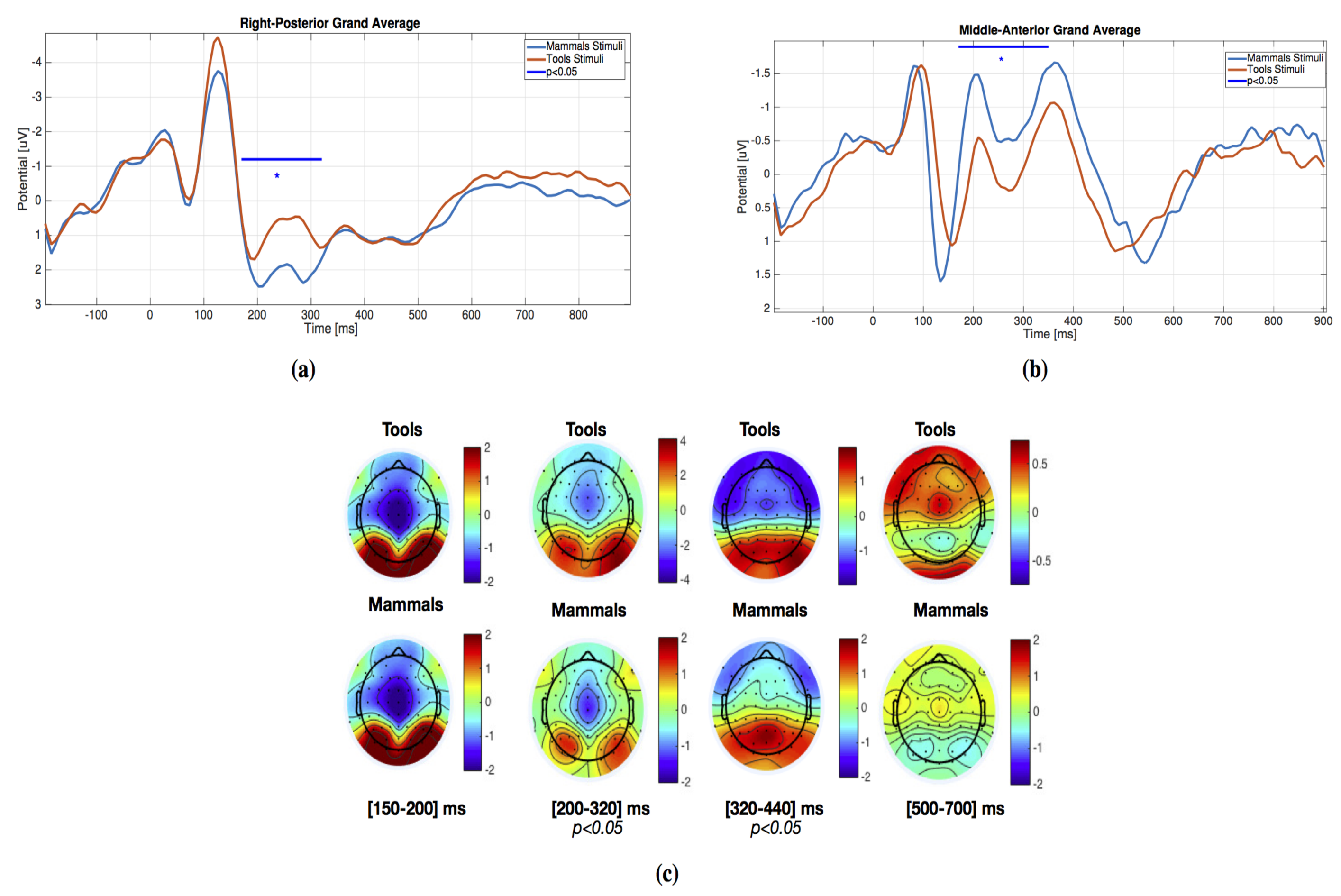

Figure 3: Time plots for right-posterior and middle anterior Figure (3a) and (3b) responses show a significant region between $[150 - 320]$ ms respectively, consistent with early visual processes thought to follow early neural responses [4]. Similarly, the scalp plots (Figure (3c)) make evident significant activations around posterior and middle-posterior regions, especially for P1 and N2 ranges (Note: refer to colorbar scale to identify the potential difference between each scalp plot).

Modalities	LOSO			5-Fold per subject			LOTO per subject		
ROIs	kNN-20	SVM R=0.1	DNN 20-10	kNN-20	SVM R=0.1	DNN 20-10	kNN-20	SVM R=0.1	DNN 20-10
Middle-frontal	0.535	0.567	0.731	0.401	0.699	0.782	0.232	0.544	0.687
Left-Anterior	0.532	0.651	0.746	0.443	0.631	0.771	0.351	0.561	0.715
Right-Anterior	0.510	0.583	0.730	0.467	0.681	0.789	0.421	0.589	0.688
Middle-Anterior	0.545	0.657	0.721	0.378	0.674	0.797	0.444	0.621	0.734
Middle-Posterior	0.578	0.527	***0.752***	0.452	0.731	***0.824***	0.501	0.666	0.751
Left-Posterior	0.612	0.611	0.744	0.534	0.743	***0.834***	0.411	0.642	***0.633***
Right-Posterior	0.624	0.626	***0.758***	0.452	0.761	***0.853***	0.455	0.611	***0.624***
Average	0.562	0.603	**0.740**	0.447	0.703	**0.807**	0.402	0.605	**0.691**
All-ROI$_s$	0.601	0.633	**0.751**	0.527	0.718	**0.838**	0.565	0.673	**0.744**

Table 1: Accuracy average results for LOSO and 5-Fold modalities specified per ROI. Bold+italics values are $p < 0.05$ based on a t-test inter-classifier comparison. Additionally, we can achieve high performance taking only $[0 - 500]$ ms ranges.

Conclusions

In this work we developed a novel pipeline for semantic decoding, outstripping the current state-of-the-art performance, thus validating the semantic-decoding process embedded in neural responses. Likewise, the performance achieved with these methodology is an open door for new Deep-Learning systems implementation, especially for semantic decoding and an extended number of semantic-classes and stimuli modalities.

References

[1] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[2] Thomas Carlson, David A Tovar, Arjen Alink, and Nikolaus Kriegeskorte. Representational dynamics of object vision: the first 1000 ms. *Journal of vision*, 13(10):1–1, 2013.

[3] Radoslaw Martin Cichy, Aditya Khosla, Dimitrios Pantazis, and Aude Oliva. Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *bioRxiv*, page 032623, 2015.

[4] Angela D Friederici and Wolf Singer. Grounding language processing on basic neurophysiological principles. *Trends in cognitive sciences*, 19(6):329–338, 2015.

[5] Nienke Meulman, Laurie A Stowe, Simone A Sprenger, Moniek Bresser, and Monika S Schmid. An erp study on l2 syntax processing: When do learners fail? *Learning a non-native language in a naturalistic environment: Insights from behavioural and neuroimaging research*, page 21, 2015.

[6] Brian Murphy, Massimo Poesio, Francesca Bovolo, Lorenzo Bruzzone, Michele Dalponte, and Heba Lakany. Eeg decoding of semantic category reveals distributed representations for single concepts. *Brain and language*, 117(1):12–22, 2011.

[7] Irina Simanova, Marcel van Gerven, Robert Oostenveld, and Peter Hagoort. Identifying object categories from event-related eeg: Toward decoding of conceptual representations. *PLoS ONE*, 5(12), 2010.

Acknowledgements

We want to acknowledge prof. Brian Murphy from Queen University Belfast (QUB) Ireland, for providing access to his data presented in [6], thus allowing us to evaluate this pipeline properly.