Classifying the mental representation of word meaning in children with Multivariate Pattern Analysis of fNIRS

Jessica Gemignani [1, 2], Laurie Bayet [3], Claire Kabdebon [4], Benjamin Blankertz [2], Kenneth R. Pugh[4], Richard N. Aslin [4]

***Abstract*— This study presents the implementation of a within-subject neural decoder, based on Support Vector Machines, and its application for the classification of distributed patterns of hemodynamic activation, measured with Functional Near Infrared Spectroscopy (fNIRS) on children, in response to meaningful and meaningless auditory stimuli. Classification accuracy nominally exceeds chance level for the majority of the participants, but fails to reach statistical significance. Future work should investigate whether individual differences in classification accuracy may relate to other characteristics of the children, such as their cognitive, speech or reading abilities.**

I. Introduction

The increasing development and use of multivariate statistical analysis techniques represents a significant advance in the field of neuroimaging. These methods have been shown to make possible the discovery of subtle effects that may remain undetected by using univariate analysis of spatially clustered regions of interest [1][2]. For this reason, their use has become widespread especially with fMRI data, and gradually also with other neuroimaging data [3]. Many studies employing multivariate pattern analysis (MVPA) techniques have confirmed that the distributed patterns of brain activity can decode many different classes of stimuli [4], but to date there is no reported use of MVPA to discriminate patterns of cerebral hemodynamic responses to purely auditory stimuli that only differ by the meaningfulness, or lack thereof, of the presented words.

Contrasting hemodynamic responses to words and non-words has the potential to identify how different brain regions contribute to lexical processing and how individual participants access words in the mental lexicon. [5]

Here we propose to reverse this logic and to use the hemodynamic pattern over different brain regions to identify this contrast in word meaning, with the ultimate goal of characterizing the lexical skills at an individual level. To this end, we present children with meaningless and meaningful stimuli, and we employ classification methods to discriminate their elicited brain patterns by means of a within-subject decoder based on the use of Support Vector Machines (SVM) that uses features drawn from the simultaneous variation of both oxygenated and deoxygenated components of the hemodynamic response measured with functional Near Infrared Spectroscopy (fNIRS).

[1] NIRx Medizintechnik GmbH, Gustav-Meyer-Allee 25, 13355 Berlin, Germany

[2] Neurotechnology Group, Technische Universität Berlin, Marchstraße 23, 10587 Berlin, Germany

[3] Laboratories of Cognitive Neuroscience, Boston Children's Hospital, 1 Autumn Street, Boston MA 02215, USA

[4] Haskins Laboratories, George Street 300, New Haven, CT 06511, USA

II. Methods

A. Participants

Seventy-two (72) healthy children (38 males, 34 females) between the ages of 3.72 and 7.76 years (M: 5.64, S: 1.05) participated in this study. Participants were recruited from New Haven and the surrounding areas of Connecticut. The experimental protocols were approved by the Yale Institutional Review Board. Participants were native English speakers.

B. Experiment

Participants were instructed to passively listen to 16 blocks of auditory stimuli, played though headphones while looking at a fixation cross that appeared on a monitor.

Each auditory condition consisted of 8 blocks, with each block consisting of the repeated sequence of one meaningful or meaningless word (from now on called "words" and "non-words" for brevity). The non-words conformed to the phonological properties of English, but had no meaningful referent . The duration in each block was 7 s and contained 6 repetitions of the same word or non-word, with 100 ms of silence between each repetition. There was a 13 s rest period between each block and the order of blocks (word vs. non-word) was randomized. The procedure is illustrated in Figure 3.

Throughout the exposure to these stimuli, participants' fNIRS signals were measured.

C. Data collection

The fNIRS recordings were performed with a Shimadzu Lab NIRS system with 20 sources x 18 detectors, resulting in 58 channels (sampling frequency 15.385 Hz, wavelengths 780, 805, 830 nm). Probes were positioned according to the scheme in Fig. 2. Sources and detectors were separated by 2.75-3 cm, depending on the size of the individual's head.

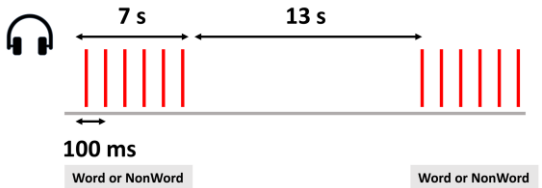

Figure 1: Scheme of the experimental design

Figure 2: Top view of the channels arrangement on the scalp

The MVPA analysis was run within-participants, therefore there was no need to precisely align the anatomical positions of the channels across participants.

D. Pre-processing

Data were pre-processed using custom scripts written in MATLAB. In particular, raw data were converted into optical density and then concentration changes of oxy- and deoxyhemoglobin (HbO and HbR) using the modified Beer-Lambert Equation ($\Delta HbO = -1{,}488 \times \Delta A780 + 0{,}5970 \times \Delta A805 + 1{,}4847 \times \Delta A830$, $\Delta HbR = 1{,}845 \times \Delta A780 - 0{,}2394 \times \Delta A805 - 1{,}0947 \times \Delta A830$, with A780, A805 and A830 being the optical absorbances at 780, 805 and 830 nm, respectively).

After conversion into HbO and HbR changes, time traces were screened to detect and correct motion artifacts. This step was performed using the Wavelet-based algorithm described in [6] and available in Homer2 [7], and applied with parameter threshold =1.5 and order N=2 Daubechies wavelet. After the motion artifacts correction, the data quality of the channels was evaluated by computing the coefficient of variation (CV) of each channel, defined as the standard deviation of the timeseries divided by its mean value. Channels having CV > 8% for either HbO or HbR were discarded from the analysis. This criterion was defined by visual inspection.

The next step was to remove from the data the contribution of global systemic effects originating in the superficial layers of scalp, dura, and peripheral vasculature and not in the cortex [8]. To this end, a spatial filter based on principal component decomposition [9] was applied to both HbO and HbR time traces.

Finally, a band-pass frequency filter was applied to the data in the range 0.01-0.2 Hz [10], to ensure that the time traces did not contain contributions from either the very low frequency systemic fluctuations, the respiration or the heart-rate oscillations. The filter was designed as a zero-phase digital FIR filter with the MATLAB function *filtfilt.*

After pre-processing, time traces were epoched in a time window of -0.5 s before each block to 15 s after each block, and baseline corrected (baseline: -0.5 s to 0). After epoching, a "channel stability" analysis was performed to determine which channels did not respond reliably across multiple blocks of stimuli. This analysis, initially introduced for fMRI analysis, has been successfully adapted to fNIRS data [11]. For each channel, irrespective of the type of stimulus (i.e., it is not a measure of discriminability of words vs. non-words), the Pearson correlation coefficients are computed

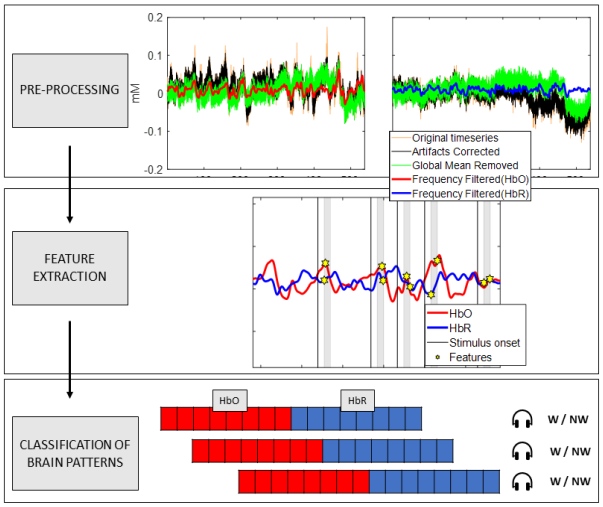

Figure 3: Scheme of the data analysis pipeline

between each possible pair of blocks, and then averaged to produce a mean stability value.

Only channels having a stability value equal or higher than the median stability value across all channels are retained for further analysis. This procedure was carried out separately on HbO and HbR time traces, and channels that resulted stable at least for one of the two hemoglobin components were retained for further analysis.

For each subject, the remaining blocks were then classified as "words" or "non-words" using the algorithm explained in the following section ("within-subject classification").

E. The algorithm

- *Extraction of features*

Features were extracted for each block and for each channel from both HbO and HbR time traces, separately. The largest peak values of HbO and HbR were computed within the time window from 5 to 10 seconds after onset of the stimulus, since it's the time interval where the peak of the hemodynamic response is most likely to occur [12].

To ensure the meaningfulness of the features in representing the hemodynamic activation in response to the task, the peak value was computed as the largest positive value for HbO and the largest negative value of HbR, because HbR is expected to have a negative deflection in response to neural activation [13], [14]. Each feature was then normalized, by removing its mean value across trials and dividing by its standard deviation.

- *Classification of Multivariate Patterns*

The obtained trials were classified using linear support vector machines implemented in MATLAB 2017b with libsvm-3.11 [15], with 4-fold cross-validation and 100 permutations. Sixteen blocks were randomly divided into 4 folds, with each fold containing 2 blocks for each class. Three folds and the remaining fold were used for training and test dataset, respectively. This procedure was repeated 100 times, and at each repetition the order of the trials was

permuted. For each subject, classification accuracies were averaged across the permutations.

- *Modeling the classification accuracy*

As a result of both the CV quality criterion and the stability analysis, both performed within-subject, the number of available channels was different across subjects. The potential effect of different number of channels on the results was therefore investigated.

To investigate the source of great variability in the resulting classification accuracies, a Mixed Effect Linear (LME) Model was fitted in MATLAB 2017b, with a random intercept for each participant, and fixed effects including Age and Number of Excluded Channels (Accuracy ~ Age + ExcludedChannels + (1|Participant)) and an Analysis of Variance was carried out on the model to test the statistical significance of the effects (DF=66).

III. Results

A. *Classification accuracy*

The overall classification results showed substantial variability across subjects, ranging from 24.5% to 85.5% (median= 54.5%, μ= 53.3%, σ = 14.7 %). Figure 4 shows the distribution of classification accuracies.

B. *Evaluation of the effects*

To assess whether the signal quality, and therefore the different number of stable channels, played a role in the variability of classification results that we observed, the distribution of number of excluded channels was partitioned into 25th, 50th and 75th percentiles (0-5, 5-10, 10-15, 15-49), and the distributions of classification accuracies within these bins were evaluated (Figure 5). The fewest subjects were eliminated when only 5-10 channels were excluded, and the resultant decoding accuracy among these subjects was the highest and least variable.

However, the ANOVA did not reveal any significant effect of either Age (p=0.810) or Number of Excluded Channels (p=0.282) on the decoding accuracy. The LME model estimated a negative coefficient for Age (ß1= -0.403, SE= 1.67) and a positive coefficient for Excluded Channels

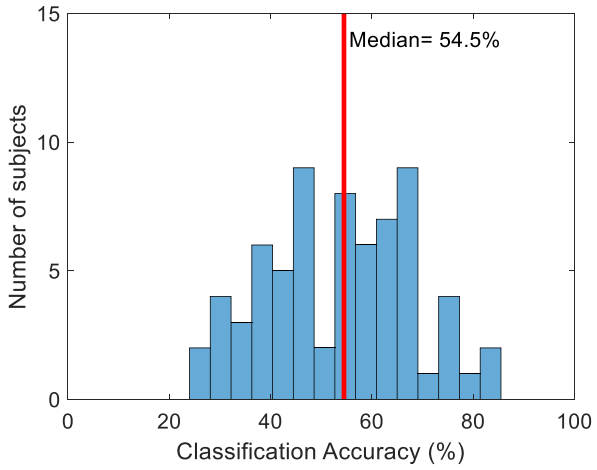

Figure 4: Distribution of overall classification accuracies (the red line indicates the median value)

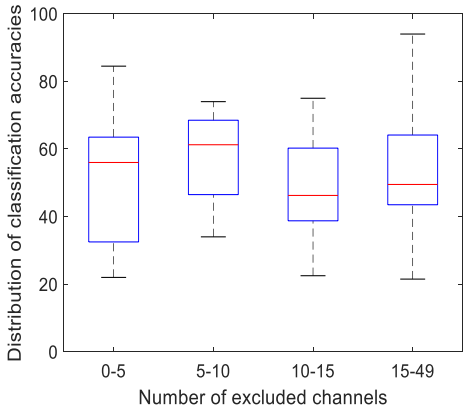

Figure 5: Distributions of classification accuracies grouped by number of unstable channels

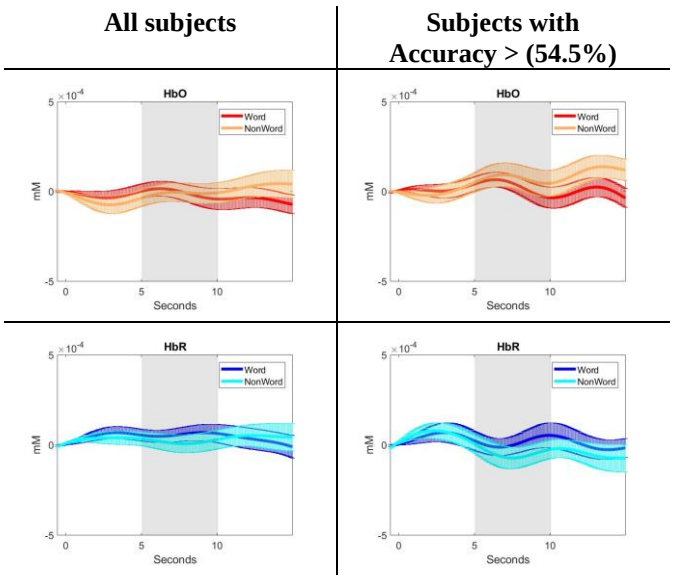

Figure 6: Grand averages over subjects and channels of the hemodynamic responses to word and non-words (top: HbO, bottom: HbR, left: average over all subjects, right: average over subjects with accuracies above the median)

(ß2= 0.176, SE= 0.16), although both of these factors played a minor role.

C. *Grand averages of the hemodynamic response to the stimuli*

Figure 6 shows the grand averages of the hemodynamic response to the two classes of stimuli, computed across the whole group (left column) and across the subjects having accuracies equal or higher than the median value (54.5%) (right column). From visual inspection, the subjects having the higher accuracies are those whose hemodynamic activation to NonWords is greater than to Words (both with HbO and HbR). To explore whether such an interaction between higher accuracies (>54.5%) and stimulus type was statistically significant, we modeled the mean values of the timetraces in the observation window (5-10sec) with a LME model. We defined 2 groups ("low" and "high" accuracy),

and used group and stimulus type as between-subject factors and channel values as within-subject factor. The test only revealed significant main effect of "Group" ($p=0.03$ for HbR and 0.08 for HbO), but not for the other effects or for the interaction group x stimulus type.

IV. Discussion

The algorithm described in this work provides evidence that distributed patterns of hemodynamic activity measured by fNIRS can, in some cases, discriminate between two classes of auditory stimuli that only differ by the meaningfulness of the presented words. Importantly, a univariate analysis revealed no differences between words and non-words, suggesting that a multivariate approach is a promising technique for decoding subtle differences in meaning.

The presented analysis is multivariate in that it combines features derived from different channels; but it is also multivariate because it combines features derived from the simultaneous variations of both hemoglobin components, which potentially makes the analysis more sensitive and less susceptible to the natural inter-subjective variability of the hemodynamic response.

The resulting classification accuracies vary greatly across the tested subjects; they are nominally above chance level for 57% of the subjects and above 60% for 36% of them, but unfortunately are not statistically above chance at the group level.

The source of this great variability is yet to be investigated; the present study excludes a statistically significant effect of signal quality or age of the participants on the results. Statistical power will likely require a larger sample size.

The method achieves accuracies below chance level for many subjects, but interestingly, by visually inspecting the time course of the responses we observe that the subjects with best accuracies show higher response to non-words than to words, although the corresponding statistical test did not reveal such interaction.

Given the wide range of ages of the participants, and so of their cognitive, reading and speech comprehension abilities, we hypothesize that the individual differences in word vs. non-word classification accuracy could be related to individual differences in cognitive, reading or speech abilities which could underlie the observed individual differences in the representation of the meaning of spoken words.

V. Conclusion

The presented method is the first of its kind to apply Support Vector Machines combining HbO and HbR as simultaneous features to classify multivariate patterns of hemodynamic activity elicited in children listening to meaningful and meaningless words. The subset of participants with the best accuracies show a higher neural activation, observable on both HbO and HbR, to non-words than to words. The next step of this work will be to investigate whether the classification accuracies are related to the ability of the children to decode the presence of meaning in words, by interrogating the relation between the accuracies and their cognitive, reading and speech comprehension skills.

Acknowledgment

The authors would like to thank the PREDICTABLE grant (Grant agreement no: 641858), "Understanding and predicting developmental language abilities and disorders in multilingual Europe" funded by the European Commission for providing financial support.

References

[1] F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns," *Neuroimage*, vol. 43, no. 1, pp. 44–58, 2008.

[2] C. Habeck and Y. Stern, "Multivariate data analysis for neuroimaging data: overview and Application to Alzheimer's disease," *Cell Biochem Biophys*, vol. 58, no. 2, pp. 53–67, 2010.

[3] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nat. Rev. Neurosci.*, vol. 7, no. 7, pp. 523–534, 2006.

[4] L. L. Emberson, B. D. Zinszer, R. D. S. Raizada, and R. N. Aslin, "Decoding the infant mind: Multivariate pattern analysis (MVPA) usingfNIRS," *PLoS One*, vol. 12, no. 4, pp. 1–23, 2017.

[5] G. Hickok and S. L. Small, *Neurobiology of language*. Academic press, 2015.

[6] B. Molavi and G. A. Dumont, "Wavelet-based motion artifact removal for functional near-infrared spectroscopy," *Physiol. Meas.*, vol. 33, no. 2, pp. 259–270, 2012.

[7] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "HomER: a review of time-series analysis methods for near- infrared spectroscopy of the brain," vol. 15, no. 10, pp. 1203–1214, 2009.

[8] I. Tachtsidis and F. Scholkmann, "False positives and false negatives in functional NIRS: 7 issues, challenges and the way forward," *J. Biomed. Opt.*

[9] X. Zhang, J. A. Noah, and J. Hirsch, "Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering," *Neurophotonics*, vol. 3, no. 1, p. 15004, 2016.

[10] N. Naseer and K.-S. Hong, "fNIRS-based brain-computer interfaces: a review," *Front. Hum. Neurosci.*, vol. 9, no. January, pp. 1–15, 2015.

[11] B. D. Zinszer, L. Bayet, L. L. Emberson, R. D. S. Raizada, and R. N. Aslin, "Decoding semantic representations from functional near-infrared spectroscopy signals," *Neurophotonics*, vol. 5, no. 1, p. 1, 2017.

[12] S. Fazli *et al.*, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage*, vol. 59, no. 1, pp. 519–529, 2012.

[13] D. T. Delpy and M. Cope, "Quantification in tissue near-infrared spectroscopy," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 352, no. 1354, pp. 649–659, 1997.

[14] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[15] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1–39, 2013.