Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-12

EVALUATING AUTOMATIC ARTIFACT CORRECTION FOR ONLINE HYPOTHESIS TESTING

A. M. Freitas[1,2], G. Sanchez[1], F. Lecaignard[1], E. Maby[1], A. B. Soares[2], J. Mattout[1]

[1] Lyon Neuroscience Research Center, Lyon, France
[2] Federal University of Uberlândia, Uberlândia, Brazil

E-mail: amanda.medeirosf@yahoo.com.br

ABSTRACT: In BCI, artifact removal remains an acute challenge. Filtering must be efficient in removing artifacts while preserving relevant features, e.g. event-related potentials (ERP) like the mismatch negativity (MMN). MMN is a prediction error signal whose modulations reflect human perceptual inference and learning. Characterizing these subtle processes requires fitting non-linear models onto single-trial data. And disentangling between alternative models is challenging because of a low signal-to-noise ratio. We evaluated four methods for online artifact removal. We mimicked online data processing using real electroencephalography (EEG) data from an auditory oddball paradigm. We compared the four approaches with standard offline analysis, in their ability to reveal (i) the MMN, (ii) the MMN modulations by the manipulation of the predictability of a sound sequence and (iii) the most likely learning mechanism at play. Artifact Subspace Reconstruction (ASR) and Empirical Mode Decomposition (EMD) were the most successful. Interestingly, they even proved more sensitive than the offline analysis, likely because they avoid rejecting trials.

INTRODUCTION

Brain-computer interfaces (BCI) measure and process brain activities for control, monitoring or rehabilitation purposes [1]. Whatever the application, an acute challenge is to extract reliable features and to translate them into meaningful information for the machine, in real time. Moreover, the challenge intensifies when brain signals are measured by non-invasive methods, typically electroencephalography (EEG).

EEG, compared with other techniques, has a poor spatial resolution and is often contaminated with electrical activities generated either by endogenous physiological sources (such as eye movements) or externally over the scalp (line noise) [1]. These artifacts, in addition to decreasing the signal quality, directly influence the classification performance of EEG-based BCIs [2] and may add to other possible reasons why some users are unable to control such BCI systems [3].

In the last few years, some approaches have been investigated to detect and remove artifacts in real-time, such as Artifact Subspace Reconstruction (ASR) [4], Fully Online and automated artifact Removal for brain-Computer interfacing (FORCe) [5], online Empirical Mode Decomposition (EMD) [6–8], and online Independent Component Analysis (ICA) [9–11]. However, none of these methods can be acknowledged as the “gold standard” yet, and to our knowledge, no research has been conducted to compare these methods with standard offline preprocessing, nor to enable online hypotheses testing for optimized cognitive neurosciences experiments [12]. Ideally, these filtering methods should provide a clean signal that contains the same relevant information as if the data would have been processed offline.

The mismatch negativity (MMN) is an automatic evoked EEG component that is typically observed during the listening of an oddball tone sequence, when subtracting the average response to deviant (rare) sounds from the average response to standard (frequent) sounds. This negative deflection occurs usually between 150 and 250 at frontal and central scalp electrodes, even in the absence of attention oriented towards the sounds [13,14].

The MMN is viewed as a prediction error signal, that is a measure of the discrepancy between the expected sensation (a standard) and the observed sensory input (a deviant). The computation of prediction errors and their resolution obey hierarchical predictive coding, where cortical processes (higher levels) send predictions to lower hierarchical levels. Whenever the current prediction fails, prediction errors are forwarded up to higher hierarchical levels, following an ascending pathway [15].

Besides, a few offline auditory-based cognitive neuroscience studies demonstrated that MMN modulations reflect an implicit learning process, such that a predictable sequence of auditory stimuli yields a

 DOI: 10.3217/978-3-85125-682-6-12

decreased MMN. In other words, the more predictable the occurrence of a deviant stimulus, the more reduced the prediction error, hence the lower the MMN amplitude [14,16]. However, revealing these learning processes requires fitting non-linear mathematical models with unknown parameters, at the single trial level and for each subject independently. It has been shown that online adaptive designs could help to optimize the selection and fitting of such learning models, at the individual level [12]. In this aim in particular, obtaining clean single trial data in real-time is highly crucial.

In this paper, we evaluate and compare the performance of ASR, online EMD, online ICA, and FORCe, for online artifact correction using real data. In our experiments, we mimic real-time data processing. Precisely, we compare the performance of those methods with offline data processing, evaluating their ability (i) to reveal a significant MMN to auditory oddball stimuli; (ii) to reveal the more subtle modulation of the MMN by the predictability of the sound sequence and (iii) to identify the right learning model based on single-trial data fitting.

MATERIALS AND METHODS

Experiment: We used the EEG dataset from Lecaignard et al. [14]. These data were obtained in 20 healthy adults (10 female, mean age: 25 $\pm$ 5 years) who passively listened to auditory oddball sequences, where the occurrence of deviant sounds was either predictable or unpredictable. We implemented the same data processing performed in this study and considered it as a standard offline approach to compare with the online artifact correction methods.

Data processing: Three of the four online artifact correction methods need to be calibrated. ASR requires a clean EEG calibration signal, while ICA requires a calibration signal that contains the artifact to be removed. EMD also requires a clean calibration signal with approximately the same window size as the subsequent epochs to be preprocessed. FORCe does not need to be calibrated. The first 30s of each block of stimulations (each subject performed 4 blocks, see [14] for more details) were used to calibrate the different methods when needed. Those initial segments were not used in the subsequent MMN analyses. These signal windows were band-pass filtered at 2-20 Hz using the inverse Fast Fourier Transform (FFT) filtering available in the EEGLAB software environment [17]. In addition, the ASR calibration signal was processed to eliminate eye blinks and samples greater than 50 μV, in other words, samples that exceeded this threshold were not used for ASR calibration. For EMD, only 1s of the 30s of the filtered signal were used.

To compute ICA we used the infomax algorithm from EEGLAB. ICA was applied to all electrodes of calibration data, and the independent components were rejected by visual inspection.

After calibration, the signal was processed over epochs starting 200ms before stimulus onset and ending 500ms after. Each epoch was also band-pass filtered at 2-20 Hz prior to processing (inverse FFT). As FORCe requires at least 1 second until it is capable of removing artifacts, specifically for this method, we used epochs starting 1.2s before stimulus onset.

The performance of ASR depends on the choice of hyper-parameters [18]. Therefore, after empirical testing, with the exception of the window size parameters (window length = 0.7s, step size = 0.5s and look-ahead = 0.2s), we used the default values.

After applying the different online artifact correction methods, we removed the last 200ms from the analysis, ensuring that the signals were not contaminated by edge artifacts introduced by the filter. Furthermore, we only used as feature of interest a specific time windows from 160 to 190 ms after stimulus onset (centered on the MMN response) and channels (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) as in Lecaignard et al. [14], which were found to be the most responsive channels. We calculated the averaged signal in that spatio-temporal region of interest, for deviant and its preceding standard stimuli in order to compute the MMN amplitude.

Models: In this part of the study, we only compared the online artifact correction methods that proved able to reveal the MMN as well as its modulation by predictability. Therefore, we applied a two-tailed *t*-test to test whether the unpredictable and predictable MMN amplitudes come from populations with unequal means, with a 5% significance level. This proved to be the case for ASR and EMD (see below). The following cognitive models were implemented for comparison:

- a null model (M0) assuming no difference between responses to standard and deviant sounds;
- static (non-learning) models including binary change detection (CD), deviant detection (DD) and linear change detection (LD) [19];
- Bayesian learning models (BL) that depend on a forgetting parameter τ, meaning that the larger τ, the wider the memory [19]. We considered four different models of this type, each corresponding to a different value of τ: 2, 6, 10, and 100.

We used the VBA (Variational Bayes Analysis) toolbox [20] for our computational model analyses. To reduce the number of inversions, the EEG signal was down-sampled from 600 Hz to 100 Hz. The VBA toolbox allows to reject specific trials in the evolution model. Therefore, for the offline method, we consider the same rejected

 DOI: 10.3217/978-3-85125-682-6-12

trials as identified in [14], and for ASR and EMD we declare the calibration period (the first 30s) of each block as rejected trials (as explained before). To find out which models outperform the static null model (M0) at a given latency, we computed the relative Free-energy (as a proxy for relative model evidence) [19].

RESULTS

The left upper panel in Fig. 1a displays the grand average responses computed offline from the identified responsive channels, for the standard, the deviant and the MMN. The right upper panel in Fig. 1a shows the MMN values computed in the window of interest (160-190 ms) for all subjects and all artifact correction methods. Except for ICA, all methods revealed a significant auditory MMN.

Fig. 1b summarizes the methods ability to reveal MMN modulation by predictability. The left lower panel shows the modulation of the MMN obtained offline. The right lower panel displays the MMN amplitude for the predictable and unpredictable conditions, as obtained with each online artifact correction method. The two-tailed t-test ($p < 0.05$) showed that, like the offline analysis, ASR and EMD proved able to reveal the significant modulation of the MMN by predictability.

Finally, we further investigated whether online artifact correction with ASR or EMD would yield the same conclusion compared to previous offline analysis, regarding Bayesian model selection of perceptual learning processes as reflected by single trial modulations of auditory evoked responses [21]. Fig. 2 shows the relative Free-energy (model evidence) of every model with respect to the null model (M0). Note that models DD and BL10 prevail compared to other models, at the latency of the MMN (between 160 and 200ms).

DISCUSSION

We compared artifact correction methods with offline data processing and investigated their abilities to reveal a significant MMN, to uncover subtle MMN modulation by a contextual manipulation of the sound sequence (deviance predictability), and to identify the best perceptual model based on single-trial data analysis.

Artifact correction: We explored four methods for online artifact correction on real data, mimicking real-time data processing.

In the course of ASR implementation, we noticed that noisy channels identified during calibration must be excluded. ASR applies Principal Component Analysis (PCA) transformation to filter the signal, therefore the noise present in the calibration data may be propagated to the other channels. Besides, as noise levels (internal and external) change throughout the experiment, re-calibration of the method is advisable. In our study, we performed the re-calibration for each new block and new EEG file, that is, every 674 stimuli (approximately 10 minutes of data).

The FORCe method was designed with the proposal of being fully automated, so no calibration period or parameter setting is required. However, the function requires that at least one second of the signal must be filtered. At first, we used the same time window (-200ms to 500ms) as for the other methods, but results proved worse than when using a 1-second-long window. It is worth mentioning that the method preserved the MMN waveform, but could not reveal a statistically significant modulation by predictability.

To implement the EMD method, since the channel cluster of interest was already well established, we first computed the average time series over these channels and then applied the correction. We chose this approach to avoid processing the intrinsic mode functions (IMFs) for all 62 channels, which would require a prohibitive time for real-time implementation. Although this could decrease the signal quality, since averaging smoothes the signal, the results showed that it did not affect the MMN response. Hence the numbers of channels must be considered when choosing this method.

ICA showed the worst performance to reveal the MMN. Probably because of the simple and rigid manner we applied ICA. Indeed, we simply apply the initially estimated mixing matrix to all subsequent trials. One could imagine updating this matrix or identifying whether a given trial needs to be filtered or not, and which components should be removed. Furthermore, the size of the time window used to decompose the signal can be enhanced to better estimate the independent components.

All methods were applied to epoched data. This strategy is well suited for ERP analysis and alleviates the computational burden.

The artifact correction methods investigated in this study were able to reveal the MMN response. However, ICA excessively smoothed the signal waveform, altering the main component of the MNN peak between 150 and 200ms.

These findings look promising as they open the way to reliable online MMN analyses, that could have far-reaching applications with patients. The MMN has indeed long been investigated to assess impaired cognitive functions in various clinical populations [22].

 DOI: 10.3217/978-3-85125-682-6-12

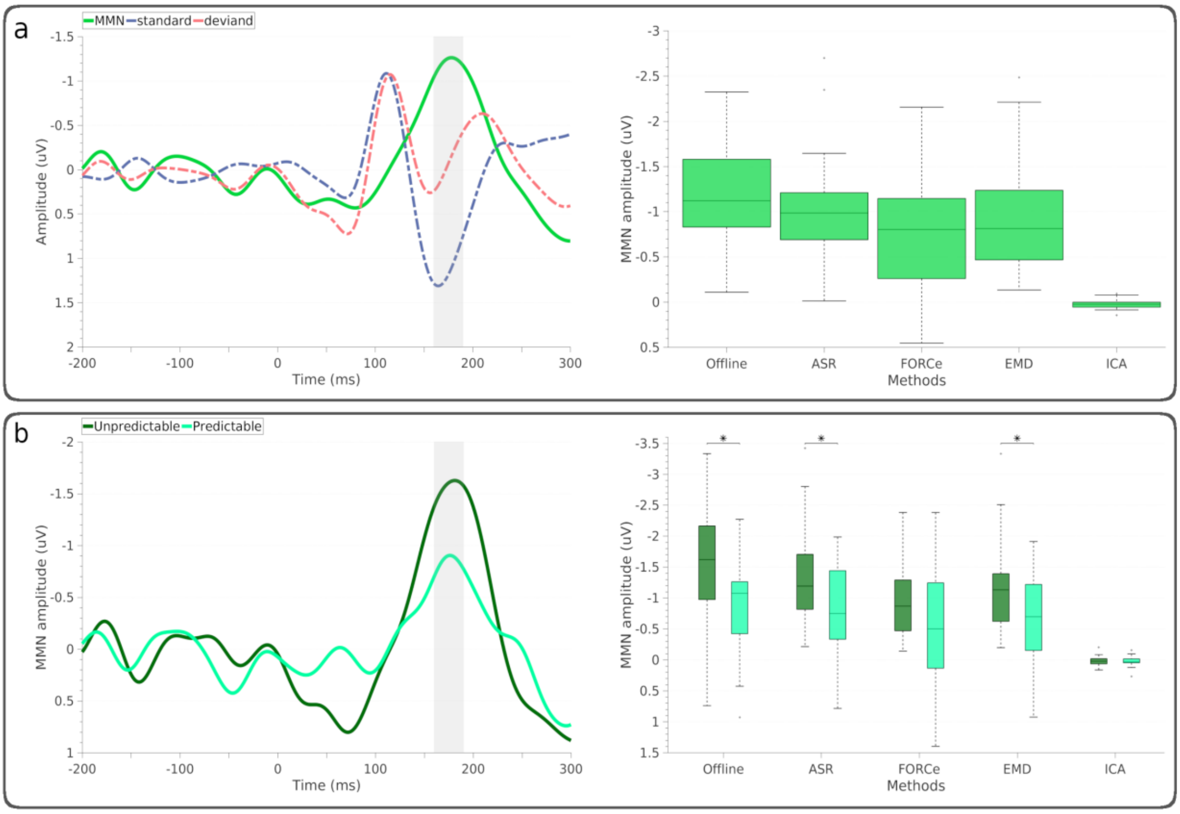

Figure 1 - Mismatch Negativity (MMN) (electrodes of interest: F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2). **(a)** Grand-average ERPs elicited by standard, deviant and their difference (MMN) obtained offline; boxplot of MMN amplitudes (averaged between 160 and 190 ms) obtained with the different online artifact corrections; **(b)** Effect of predictability on the MMN (offline method); MMN amplitudes for each condition (predictable and unpredictable) and each artifact correction approach; the asterisks indicate statistical significance of the modulation by predictability ($p < 0.05$, two-tailed *t*-test).

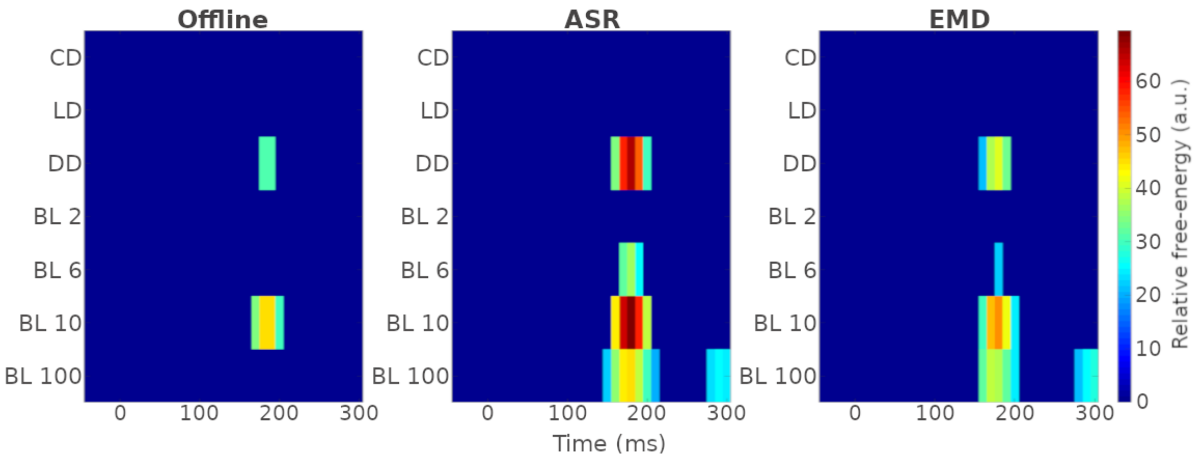

Figure 2 - Relative Free-energy maps. Relative Free-energy or (log) Bayes-factor values obtained offline (left panel), with ASR (middle panel) and EMD (right panel), respectively. In each panel, the relative Free-energy is given for each alternative model relative to the null model (y-axis) and for each peri-stimulus time sample in ms (x-axis). Values greater than 20 indicate significant evidence in favor of the alternative model.

 DOI: 10.3217/978-3-85125-682-6-12

Predictability effect: After this first analysis, we investigated the impact of artifact correction on MMN modulation. As discussed in other studies [14, 23, 24], predictability of the occurrence of a deviant sound yields a decrease in prediction error, hence a decrease in the MMN amplitude. In other words, the greater the sound predictability, the smaller the MMN.

Results depicted in Fig. 1b shows that FORCe and ICA hardly revealed the MMN modulation. Although the MMN in the predictable condition was smaller when using the FORCe approach, it was not significantly different from the one in the unpredictable condition. Whereas for the ICA approach, the estimated amplitudes were the same in the two conditions.

Models: The last part of our study was implemented to analyze if any learning or non-learning models, based on sensor-level signals, could explain the trial-by-trial variations in the ERP signals. At the latency of the MMN (150 to 200ms), the offline method presented favorable results for models DD and BL10, but the DD model took longer to stand out and its window of significance proved shorter. The Free-energy map for ASR and EMD reveal similar latencies of significance. However, the ASR approach yielded slightly longer time windows and presented larger values compared to the other methods. This suggests that ASR was not only able to appropriately correct for artifacts but in doing so, it also avoided rejecting trials and led a more sensitive analysis, better suited to reveal the minutiae of auditory processing dynamics.

CONCLUSION

Our results suggest that ASR and EMD can perform accurate online artifact detection and correction, as they were able to reproduce the results obtained at the group level with classical offline analysis. Interestingly, in terms of model comparison, we also note that the mimicked online approaches based on ASR and EMD yielded more sensitive results than the reference offline approach. However, ASR is computationally more efficient than EMD. Overall, ASR offers great potential for real-time applications, in particular those that would exploit adaptive designs in order to optimize hypothesis testing or clinical diagnosis at the individual level [12].

REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil. Brain Computer Interfaces, a Review. *Sensors*, vol. 12, no. 2, pp. 1211–1279, Feb. 2012.

[2] F. Lotte *et al.*. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.*, vol. 15, no. 3, p. 031005, Apr. 2018.

[3] D. Tan and A. Nijholt. Brain-Computer Interfaces and Human-Computer Interaction. Brain-Computer Interfaces. Springer London, 2010, pp. 3–19.

[4] T. R. Mullen *et al.*. Real-time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. IEEE Trans. Biomed. Eng., vol. 62, no. 11, pp. 2553–2567, Nov. 2015.

[5] I. Daly, R. Scherer, M. Billinger, and G. Müller-Putz. FORCe: Fully Online and Automated Artifact Removal for Brain-Computer Interfacing. IEEE Trans. Neural Syst. Rehabil. Eng., vol. 23, no. 5, pp. 725–736, Sep. 2015.

[6] R. Fontugne, P. Borgnat, and P. Flandrin. Online Empirical Mode Decomposition. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 4306–4310.

[7] A. O. Andrade, S. Nasuto, P. Kyberd, C. M. Sweeney-Reed, and F. R. Van Kanijn. EMG signal filtering based on Empirical Mode Decomposition. Biomed. Signal Process. Control, vol. 1, no. 1, pp. 44–55, Jan. 2006.

[8] Huang Norden E. *et al.*. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[9] R. N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalogr. Clin. Neurophysiol., vol. 103, no. 3, pp. 395–404, Sep. 1997.

[10] A. K. Barros, A. Mansour, and N. Ohnishi. Removing artifacts from electrocardiographic signals using independent components analysis. Neurocomputing, vol. 22, no. 1, pp. 173–186, Nov. 1998.

[11] M. Milanesi, N. Martini, N. Vanello, V. Positano, M. F. Santarelli, and L. Landini. Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals. Med. Biol. Eng. Comput., vol. 46, no. 3, pp. 251–261, Mar. 2008.

[12] G. Sanchez, F. Lecaignard, A. Otman, E. Maby, and J. Mattout. Active SAmpling Protocol (ASAP) to Optimize Individual Neurocognitive Hypothesis Testing: A BCI-Inspired Dynamic Experimental Design. Front. Hum. Neurosci., vol. 10, 2016.

[13] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clin. Neurophysiol., vol. 118, no. 12, pp. 2544–2590, Dec. 2007.

[14] F. Lecaignard, O. Bertrand, G. Gimenez, J. Mattout, and A. Caclin. Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory

 DOI: 10.3217/978-3-85125-682-6-12

hierarchy. Front. Hum. Neurosci., vol. 9, 2015.

[15] G. G. Parras, J. Nieto-Diego, G. V. Carbajal, C. Valdés-Baizabal, C. Escera, and M. S. Malmierca. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat. Commun., vol. 8, no. 1, p. 2148, Dec. 2017.

[16] V. B. Perez *et al.*. Mismatch Negativity is a Sensitive and Predictive Biomarker of Perceptual Learning During Auditory Cognitive Training in Schizophrenia. Neuropsychopharmacology, vol. 42, no. 11, pp. 2206–2213, Oct. 2017.

[17] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[18] C. Chang, S. Hsu, L. Pion-Tonachini, and T. Jung. Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 1242–1245.

[19] D. Ostwald, B. Spitzer, M. Guggenmos, T. T. Schmidt, S. J. Kiebel, and F. Blankenburg. Evidence for neural encoding of Bayesian surprise in human somatosensation. NeuroImage, vol. 62, no. 1, pp. 177–188, Aug. 2012.

[20] J. Daunizeau, V. Adam, and L. Rigoux. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLOS Comput. Biol., vol. 10, no. 1, p. e1003441, Jan. 2014.

[21] F. Lecaignard, O. Bertrand, A. Caclin, and J. Mattout. Evidence for implicit and adaptive deployment of precision weighting during passive listening: a simultaneous EEG/MEG study. Neuroscience, preprint, Dec. 2018.

[22] R. Näätänen, E. S. Sussman, D. Salisbury, and V. L. Shafer. Mismatch Negativity (MMN) as an Index of Cognitive Dysfunction. Brain Topogr., vol. 27, no. 4, pp. 451–466, Jul. 2014.

[23] C. Wacongne, J.-P. Changeux, and S. Dehaene. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J. Neurosci., vol. 32, no. 11, pp. 3665–3678, Mar. 2012.

[24] I. Winkler and I. Czigler. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int. J. Psychophysiol., vol. 83, no. 2, pp. 132–143, Feb. 2012.