

Predictive Processing and temporal predictability in visual domain.

Sanjeev Nara ¹ Mikel Lizarazu ¹ Craig Richter ¹ Diana Dima ⁴ Radoslaw Martin Cichy ³ Mathieu Bourguignon ^{1,2} Nicola Molinaro ¹

1. Basque Center for Cognition, Brain and Language , San Sebastian, Spain
2. Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium
3. Department of Education and Psychology, Free University Berlin, Berlin, Germany
4. John Hopkins University, Baltimore , USA

Introduction :

Predictive processing is framework which explains how does the brain interacts with the external environment through sensory modalities. The core idea behind all the predictive processing theories is that the brain predicts about the environment around it and develops a generative model. The model is constantly updated through the comparison of predicted and actual stimuli. This process follows a cortical hierarchy (processing hierarchy), the higher order brain areas makes a prediction (top – down descending activation) and the lower order brain areas (bottom up ascending activations) gets the actual stimuli input. The comparison between both the actual and predicted input is termed as prediction error and this error signal is used to update the internal model of environment around us [1].

This framework has also been termed by different names like predictive coding, hierarchal predictive coding etc but the key elements remain the same “*Prediction*” and “*Prediction errors*” [2], [3]. It is well supported in literature that predictive processing framework has been able to account for underlying cognitive processes both in humans as well as animals [4]. Predictive processing accounts for visual perception [5][6] and auditory perception [7], speech perception[8], language [9] and music [10]. The framework is also used in explaining the abnormal functioning in neural disorders like autism [11], psychosis [12] and several others.

In this hierarchical framework, the input stimuli to the lower order brain areas may have several features. A visual stimulus may have features like edges,

orientations, color, shape and other properties, whereas the audio signal might have features like volume, amplitude, frequency, pitch etc. These features form the basis of perception and are related to the identity of the stimuli often contain the “WHAT” information. But in real life situation the timing of the input stimuli also matters a lot. You may see a car coming toward you on the road, even after having all the “WHAT” information features, it is really important to know the “WHEN” information to avoid a collision. The same example holds valid in auditory domain i.e how fast or slow you listen to the input shapes your perception. This WHEN information has gained less interest compared to the WHAT information. These WHAT and WHEN features interact with each other and shape the human perception so there is also a need to understand the interaction of WHAT and WHEN in human perception. There have been attempts by Faidella et al [13] to understand this interaction in auditory domain, but the limitation of the EEG modality (restricted to source level inferences only) and experimental design did not shed proper light on this interaction. There were also attempts by Aukstulewicz et al [14] to study this interaction using ECOG data, but the less no of participants used and exclusion of an important factor i.e Attention, failed to investigate the proper interaction of this What and When.

Attention and Prediction both are top down processes, both affect human perception. Theories on prediction relate to neural (mental) models inferring about (present or future) sensory or action-related information, whereas theories of attention are about the control of information flow underlying perception and action. Both concepts are related and not always clearly distinguishable. [15] Whereas there have been a lot of evidence about the feature specific attentional abilities which means the focussed and divided attention can be dissociated [16][17][18]. This gives us motivation to investigate attention and prediction by designing an orthogonal task, where attention is diverted to one feature of the stimuli whereas the prediction is related to another feature. This design should analyse the prediction effects in less/diverted focus of attention.

Returning back to the WHEN information, this information about the stimuli might be either predictable or unpredictable (temporal predictability in general). Most of the experiments in predictive processing research have studied prediction effects with just two components i.e a cue and a target. This has been reported as match and mismatch prediction effects based on match/mismatch of cue and target. To investigate the effect of temporal predictability on prediction, there should be longer sequences of cue's which

will generate stronger predictions, helps to understand how does the temporable predictability affect the development of predictions.

In this experiment we focus on visual domain, interaction of WHAT and WHEN information and investigation the prediction effects in less focus of attention. We manipulated the attention of the participants to one of feature of the stimuli (cpd : cycles per degree of a gabor patch) and the prediction to the another feature (orientation of the gabor patch) in a series of previous stimuli in a sequence of four gabors. The feature having focussed attention is present at the Target. Here we plan to measure the effect of temporal predictability on the four gabor patches which are in less focus of attention.

The presentation of gabors may enhance the neural response or may also suppress the neural response based on temporal predictability. This kind of suppression or enhancement has been studies through different names like Neural adaptation, repeation suppression, repeatition enhancement, mnemonic filtering, decremental response and Neural priming [19]. On one side it is possible that temporal predictability could enhance repeatition suppression effect as there are two sources of information about the stimuli i.e timing and orientation. On the other side it is possible that the predictions are enhanced when the timing information is unpredictable as the system increase the reliance on the internal model to reduce the uncertainty of the external stimuli. We do so through conventional evoked analysis approach and advanced analysis like source reconstruction.

With the access to higher computational power and advanced machine learning methods, multivariate pattern analysis has gained a lot of interest in recent years. It is now possible to decode the orientation, position, identity, angle, seen and unseen information from EEG and MEG signals at every millisecond or every sample [20][21], [22]. I will be interesting to decode the orientation of each presented gabors during less focus of attention, representing the developing predictions.

Methods :

Participants :

Seventeen (xx Females) took part in the present study with mean age xx, std- yy and age range xx- yy years. The ethical committee and scientific committee of Basque Center on Cognition, Brain and Language (BCBL) approved the experiment (following the principles of the Declaration of Helsinki) and each participant signed an informed consent form. Each participant was paid 10 € per hour for their valuable contribution. The participants were recruited from BCBL web participa and all the participants were free from any neurological and psychological disorders. All the participants had normal or corrected to normal vision.

Experiment Design :

To address the experimental questions, we developed an experimental paradigm in which participants were presented with simple visual stimuli (gabor patches) that could vary in a number of dimensions. Four gabor patches were presented for 200 milliseconds sequentially (with the inter-stimulus interval of 200 ms) and constitute the contextual information (we call them entrainers). After a longer interval (inter-stimulus interval of 800 ms), a fifth gabor patch (target) was presented and participants had to perform a task on this. The entrainers had an intermediate number of cycles per degree (cpd) of visual angle: participants had to detect if the target had an either higher or a lower cpd with respect to the entrainers. Prediction, however, was focused on a different feature of the stimuli. In fact, the gabor orientation always varied across the five visual stimuli (entrainers and target) but could either vary randomly or change gradually the orientation (clockwise or counterclockwise). In the latter case, the orientation of the target was predictable. This design nicely drives the attention of the participants on the cpd feature of the gabor patches, while manipulating the orientation for estimating the predictive processes that are supposed to be automatic. The design includes a temporal prediction dimension as well: the time gap (inter-stimulus interval) before the target gabor could be predictable (fixed time gap) or could be unpredictable (random time gap). The inclusion of this dimension makes the design predictable and non-predictable in the two dimensions (predicting what and predicting when) :

	Predictable gabor orientation	Unpredictable gabor orientation
Predictable time gap	Predictable Orientation, Predictable Timing (WHAT + WHEN)	Unpredictable Orientation, Predictable Timing (WHEN Only)
Unpredictable time gap	Predictable Orientation, unpredictable Timing (WHAT Only)	Unpredictable Orientation, Unpredictable Timing (Random / None)

Table 1: Experimental design

The table above shows the four different conditions . The experimental paradigm used to generate the above mentioned four conditions is shown in figure 2:

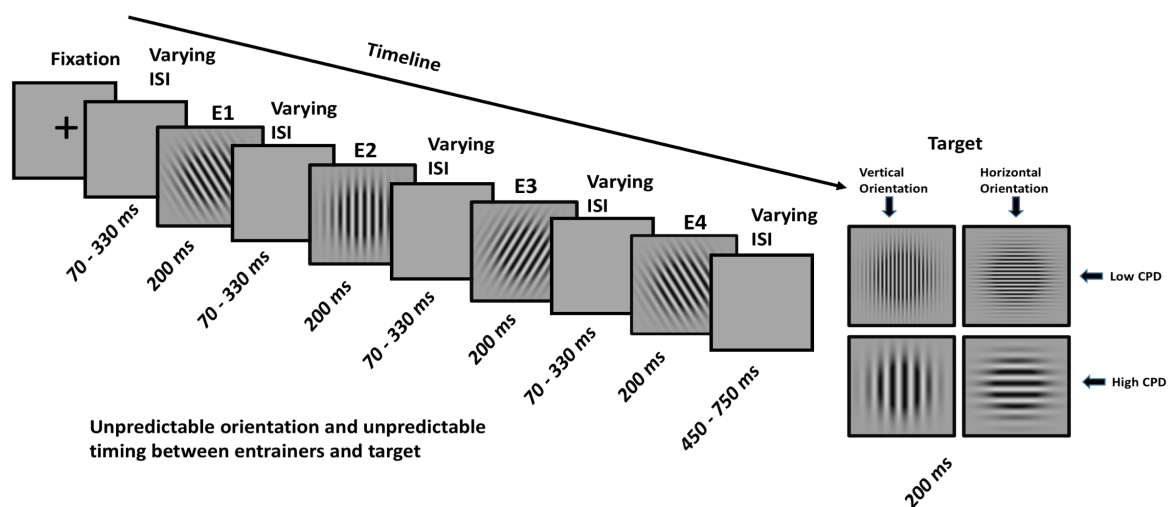


Figure 2: Experimental Stimuli

The experimental paradigm for each trial consists of 4 entrainers displayed as E1, E2, E3 and E4, followed by the target gabor patch (Target). The time gap between entrainers is represented as G1, G2, and G3, and the time gap between entrainer 4 and target gabor patch is displayed as G4. In Condition 1 (Figure 2a) the orientation of target gabor is predictable (gabor orientation, displayed with blue arrows, is scaled across entrainers), as well as the timing between the entrainers (time gap is fixed, displayed by the square wave in the figure). In condition 2 (Figure 2b) the orientation of target gabor is not predictable (gabor orientation, displayed with blue arrows, is random across entrainers) whereas the timing between the entrainers is predictable (time gap is fixed, displayed by the square wave in the figure). In Condition 3 (Figure 2c) the orientation of the target is

predictable (shown with blue arrows across the entrainers in scaled fashion) but the timing across entrainers (G1, G2, G3) and before target (G4) is unpredictable (shown by the jittering square wave). In Condition 4 (Figure 2d) both the orientation and the timing of the target are unpredictable.

Data Acquisition :

MEG data were acquired at MEG facility available at Basque Center on Cognition, brain and Language (BCBL, San Sebastian, Spain) in a magnetically shielded room using the whole-scalp MEG system (Elekta NeuromagR®, Helsinki, Finland). The MEG system is equipped with 306 sensors distributed around the participant's head. The sensors are arranged in triplet configuration of one magnetometer and two orthogonal planar gradiometers. To keep a track of head movement five (05) magnetic head position indicator (HPI) coils were also placed around the head. The initial location of the hpi coils and the three reference points i.e nasion, right auricular point and left auricular point was obtained by a 3D digitizer (Polhemus Fastrak, Colchester, VA, USA). This initial and reference point information is crucial in head motion correction during max filtering. Three hundred (300) additional points of the scalp of participant's head were also obtained which helps in better coregistration of the MRI T1 images of each participant acquired at the MRI facility available at BCBL. The MRI data were acquired using a three tesla (3T) MRI machine (Siemens Medical System, Erlangen, Germany) EPI xx sequence and xx mm voxel size. The MEG data was acquired at 1kHz sampling rate with bandpass filter (xx order) . in range 0.1 to 330 Hz. Cardiac artifacts (ECG) and ocular artifacts (EOG, both blinks and saccades) were also recorded during the acquisition.

Data preprocessing :

The initial preprocessing were performed using maxfilter software (Maxfilter 2.2). The Spatiotemporal signal space separation (tsss) along with movement correction was applied to the data. The bad channels were interpolated from neighbouring sensors with algorithms implemented in maxfilter software. The further preprocessing was performed using Fieldtrip toolbox [23] and custom

scripts written in Matlab 2014b (Mathworks). The data was corrected for jump artifacts and muscle artifacts. The ocular and cardiac artifacts were automatically removed using ICA (runica) and coherence based method (add reference for tutorial).

Evoked Analysis :

The data were bandpassed in the range 0.5 - 45 hz for doing an evoked analysis. The gradiometers were combined resulting 102 channels which were used for subsequent analysis. The trials were combined across the corresponding conditions. The data was time locked to target (T) for every condition to investigate the onsets of all the conditions . **Figure xx** shows the mean activity over all the participants in all the conditions. It is evident that the evoked activity of entrainers is lost when the data is time locked to target. so the data for all the entrainers and target was time locked individually. **Figure XX** shows the data time locked to entrainers

Effect of temporal predictions on visual evoked Potentials :

To investigate the effect of temporal predictions on visual evoked potentials, data from each entrainer and target was selected between 70 - 150 msec after the onset of trigger. The peak amplitude in this time range was computed for every condition and all the participants. **Figure XX** shows the peak evoked potential across in conditions with and without temporal conditions.

To assess the statistical difference in the peak amplitudes across conditions, the peaks in condition What + When (temporally predictable) was compared with When condition. Ttest ($\alpha = 0.001$) when applied to the peaks shows there is significant difference in the peaks in both conditions which emerge at Entrainer 2. The same statistical comparison applied to condition What only and random/No prediction (temporally unpredictable), the statistical significance occurs at Entrainer 3. This provides evidence that the inclusion of temporal predictions facilitates the visual processing of input.

Cluster Based permutation test on Evoked Potentials :

The analysis on the peak amplitudes informs us that there is a significant difference across predictable and unpredictable conditions. To investigate which are the brain regions on sensor space that contribute to this difference, cluster based permutation test was applied to the evoked activity at each entrainer. **Figure XX** shows the cluster based results on every entrainer and the target.

Source Reconstruction :

Source reconstruction was performed on evoked data to find out the brain areas generating the differential effect on sensor space. The source reconstruction pipeline used here has also been reported in Monsalve et al [24]. linearly constrained minimum variance beamformer (LCMVB)[25] approach was used for creating inverse operator. T1 weighted high resolution images from 14 participants were parcellated using freesurfer software[26]. The MEG and MRI coordinates were co-registered based on three fiducial points and points on brain scalp acquired during the head digitization. One shell realistic head model were created from the parcellated MRI data. The forward model was computed for three orthogonal source orientations, placed on a 5 mm grid covering the whole brain using MNE suite [27] (Martinos Centre for Biomedical Imaging). Each source was then reduced to its two principal components of highest singular value, which closely correspond to sources tangential to the skull. Both planar gradiometers and magnetometers were used for inverse modeling after dividing

each sensor signal (and the corresponding forward-model coefficients) by its noise standard deviation. The noise variance was estimated from the 400-ms baseline prior to the cross fixation in every trial.

A-priori time windows based in sensor space data was selected for each entrainer (85 – 125 msec post presentation of the gabor). LCMV inverse solution was used to project the ERF data into source space, using a noise co-variance matrix estimated from a 400 ms before the cross fixation. Source power in the selected time window was averaged and normalized by the power in the baseline. For computing the group level analysis, the individual MRIs were mapped to the standard Montreal Neurological Institute (MNI) brain through a non-linear transformation using the spatial-normalization algorithm implemented in Statistical Parametric Mapping (SPM8) [28], and the ensuing spatial transformations were applied to individual maps. We identified the coordinates of local maxima (sets of contiguous voxels displaying higher power than all other neighboring voxels).

MVPA:

To decode the feature specificity (i.e orientation of gabor), the data was segmented at every entrainer in time range -100 msec till 350 post presentation of gabor for Entrainer 1 , 2 and 3. Since the time gap between target and Entrainer was more compared to earlier entrainers, the data was segmented from -100 msec till 600 msec post presentation of the gabor patch. The data was classified using a linear support vector machine (SVM) classifier with L2 regularization and a box constraint of $c=1$. The classifier was implemented in Matlab using LibLinear [29] and the Statistics and Machine Learning Toolbox (Mathworks, Inc.). We performed a binary classification on orientation of gabor relative to the target. The label to the two classes i.e horizontal or vertical were derived from the target orientation, if the target orientation was horizontal all the preceding orientation in corresponding condition were labelled as horizontal and the other way around.

To reduce the computational load the data was down-sampled to 200 hz. Pseudo trials were generated to improve the SNR by averaging the trials in a set of 5 for each class [30]. This pseudo trial generation was repeated 100 times to generate the trials with higher signal to noise ratio. The data was then randomly partitioned using k-fold cross validation ($k=5$). The classifier was trained on 4 folds and tested on 1 fold and this process was repeated for all the folds. This random assignment was performed 5 times to remove any data biasness. To improve data quality, we performed multivariate noise normalization[31]. The time-resolved error covariance between sensors was calculated based on the covariance matrix (Σ) of the training set (X) and used to normalize both the training and test sets, in order to down weight MEG channels with higher noise levels.

Results :

Discussion :

Acknowledgement :

references :

- [1] G. B. Keller and T. D. Mrsic-Flogel, “Predictive Processing: A Canonical Cortical Computation,” *Neuron*, vol. 100, no. 2, pp. 424–435, 2018.
- [2] K. Friston, “A theory of cortical responses,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 360, no. 1456, pp. 815–836, 2005.
- [3] R. P. N. Rao and D. H. Ballard, “Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects,” *Nat. Neurosci.*, vol. 2, no. 1, pp. 79–87, 1999.
- [4] C. G. Richter, W. H. Thompson, C. A. Bosman, and P. Fries, “Top-down beta enhances bottom-up gamma,” *J. Neurosci.*, vol. 37, no. 28, pp. 6698–6711, 2017.
- [5] G. Stefanics, G. Stefanics, J. Kremláček, and I. Czigler, “Visual mismatch negativity: A predictive coding view,” *Front. Hum. Neurosci.*, vol. 8, no.

- September, pp. 1–19, 2014.
- [6] L. H. Arnal and A. L. Giraud, “Cortical oscillations and sensory predictions,” *Trends Cogn. Sci.*, vol. 16, no. 7, pp. 390–398, 2012.
 - [7] M. Heilbron and M. Chait, “Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex?,” *Neuroscience*, vol. 389, pp. 54–73, 2018.
 - [8] M. Hakonen, P. J. C. May, I. P. Jääskeläinen, E. Jokinen, M. Sams, and H. Tiitinen, “Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates,” *Brain Behav.*, vol. 7, no. 9, pp. 1–15, 2017.
 - [9] Gina R. Kuperberg and T. F. Jaeger, “What do we mean by prediction in language comprehension?,” *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2017.
 - [10] S. Koelsch, P. Vuust, and K. Friston, “Predictive Processes and the Peculiar Case of Music,” *Trends Cogn. Sci.*, vol. 23, no. 1, pp. 63–77, 2019.
 - [11] P. Sinha *et al.*, “Autism as a disorder of prediction,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 42, pp. 15220–15225, 2014.
 - [12] P. Sterzer *et al.*, “The Predictive Coding Account of Psychosis,” *Biol. Psychiatry*, vol. 84, no. 9, pp. 634–643, 2018.
 - [13] J. Costa-Faidella, T. Baldeweg, S. Grimm, and C. Escera, “Interactions between ‘what’ and ‘when’ in the auditory system: Temporal predictability enhances repetition suppression,” *J. Neurosci.*, vol. 31, no. 50, pp. 18590–18597, 2011.
 - [14] R. Auksztulewicz *et al.*, “Not all predictions are equal: ‘what’ and ‘when’ predictions modulate activity in auditory cortex through different mechanisms,” *J. Neurosci.*, vol. 38, no. 40, pp. 8680–8693, 2018.
 - [15] E. B. V All and W. James, “Bridging prediction and attention in current research on perception and action,” vol. 1626, pp. 1–13, 2015.
 - [16] X. V. Wyart, N. E. Myers, and C. Summerfield, “Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention,” vol. 35, no. 8, pp. 3485–3498, 2015.
 - [17] C. Summerfield and J. A. Mangels, “Dissociable Neural Mechanisms for Encoding Predictable and Unpredictable Events,” pp. 1120–1132, 2006.
 - [18] J. Jiang, C. Summerfield, and T. Egner, “Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain,” vol. 33, no. 47, pp. 18438–18447, 2013.
 - [19] K. Grill-Spector, R. Henson, and A. Martin, “Repetition and the brain: Neural models of stimulus-specific effects,” *Trends Cogn. Sci.*, vol. 10, no. 1, pp. 14–23, 2006.
 - [20] J. R. King, N. Pescetelli, and S. Dehaene, “Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information,” *Neuron*, vol. 92, no. 5, pp. 1122–1134, 2016.

- [21] D. Pantazis, M. Fang, S. Qin, Y. Mohsenzadeh, Q. Li, and R. M. Cichy, "Decoding the orientation of contrast edges from MEG evoked and induced responses," *Neuroimage*, vol. 180, no. July, pp. 267–279, 2018.
- [22] R. M. Cichy, D. Pantazis, and A. Oliva, "Resolving human object recognition in space and time," *Nat. Neurosci.*, vol. 17, no. 3, pp. 455–462, 2014.
- [23] R. Oostenveld, P. Fries, E. Maris, and J. M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Comput. Intell. Neurosci.*, vol. 2011, 2011.
- [24] I. F. Monsalve, M. Bourguignon, and N. Molinaro, "Theta oscillations mediate pre-activation of highly expected word initial phonemes," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018.
- [25] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Transactions on," *Biomed. Eng. (NY).*, vol. 44, no. 9, pp. 867–880, 1997.
- [26] A. Dale and M. Sereno, "Improved localization of cortical surface reconstrucion: A linear approach," *J. Cogn. Neurosci.*, vol. 5, pp. 162–176, 1993.
- [27] A. Gramfort *et al.*, "MNE software for processing MEG and EEG data," *Neuroimage*, vol. 86, pp. 446–460, 2014.
- [28] K. J. Friston, A. P. Holmes, K. J. Worsley, J. -P Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Hum. Brain Mapp.*, vol. 2, no. 4, pp. 189–210, 1994.
- [29] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, no. 2008, pp. 1871–1874, 2008.
- [30] D. C. Dima and K. D. Singh, "Dynamic representations of faces in the human ventral visual stream link visual features to behaviour," 2018.
- [31] M. Guggenmos, P. Sterzer, and R. M. Cichy, "Multivariate pattern analysis for MEG: A comparison of dissimilarity measures," *Neuroimage*, vol. 173, no. February, pp. 434–447, 2018.