# Single-Trial Inference on Visual Attention

Mads Dyrholm, Søren Kyllingsbæk, Signe Vangkilde, Thomas Habekost and Claus Bundesen

*Center for Visual Cognition, Department of Psychology, University of Copenhagen, Denmark*

**Abstract.** In this paper we take a step towards single-trial behavioral modeling within a Theory of Visual Attention (TVA). In selective attention tasks, such as the Partial Report paradigm, the subject is asked to ignore distractors and only report stimuli that belong to the target class. Nothing about a distractor is observed directly in the subject's overt behavior, hence behavioral modeling of such trials involves out-marginalizing the variables that represent the distractors' influence on behavior. In this paper we derive equations for inferring a latent representation of the distractors on a Partial Report trial. This result retrodicts a latent attentional state of the subject using the observed response from that particular trial and thus differs from other predictions made with TVA which are based on expected values of observed variables. We show an example of the result in single-trial analysis of an occipital EEG component.

**Keywords:** Single-trial, EEG, latent variables, visual attention
**PACS:** 87.19.le; 87.19.lt; 87.85.D-

## INTRODUCTION

The implicit goal in parametric modeling of behavior is to summarize a set of observations by means of a smaller set of parameters whereas the typical explicit goal is to draw comparisons between parameter estimates of different conditions. The number of model parameters grow with the number of compared conditions as opposed to the number of observations, and thus a mismatch between model parameters and single-trial ambitions exist: one cannot estimate a set of parameters for each trial individually. However, as is well known, human cognition is a stochastic process and even under identical experimental conditions no two trials produce the exact same cognitive response due to uncontrollable fluctuations in the subject's mental state (state of arousal, readiness, etc.). Unobserved (latent) variables representing such mental fluctuations *may* depend on physiological measurements such as EEG [1, 2] and pupil dilation [3], but as we will show they can also be based solely on behavioral data in combination with a parametric model. We illustrate this principle within A Theory of Visual Attention (TVA) [4] by inferring latent variables that represent the distractors in a Partial Report task. These latent variables are inferred per trial, conditioned on the observed response of the subject, i.e. retrodictions which differ from general model predictions which predict behavior based solely on experimental conditions. These purely behavioral inferences could prove useful for single-trial analysis of functional neuroimaging data.

Mathematical models are used extensively in functional neuroimaging to separate random signals from random noise. This requires assumptions about both signal and noise. A typical assumption in neuroimaging modalities such as EEG and fMRI is that the neural signal has a reproducible portion which can be separated from the noise by averaging methods (producing smooth looking ERPs in the case of EEG). To link single-trial fluctuations in EEG with behavior, one needs behavioral variables that fluctuate within trials of the same type. We show an example of inference of such variables for explaining event-related modulation of oscillatory power in an occipital EEG component.

## BEHAVIORAL MODELING WITH TVA

The Theory of Visual Attention (TVA) by Bundesen [4] offers a parametric model of a subject's behavior when asked to categorize elements of a brief display. A subject is assumed to perceive a particular categorization if and only if that categorization reaches Visual Short Term Memory (VSTM) before the stimulus display disappears. All elements in a display compete in a stochastic race for a position in VSTM: the time it takes for an object to finish is assumed to be exponentially distributed. The VSTM is assumed to have a limited capacity in the sense that it can only hold categorizations of a finite number of elements at any given time. Let $K$ denote the VSTM capacity parameter. Furthermore, let the parameter $t_0$ denote the temporal threshold for visual perception, a temporal constant which must be exceeded by the display duration in order for any categorization to take place. Assuming that the capacity limitation of VSTM is not an issue, for example if the number of displayed elements does not exceed the VSTM capacity, the probability that an element $x$ reaches the VSTM on a display $S$ of duration $t$ is given by

$$\mathrm{p_E}(\, x \mid K \geq |S|,\, t > t_0\,) = 1 - \exp(-v_x[t-t_0]) \tag{1}$$

where $v_x$ is the 'processing rate' with which the particular categorization takes place. The VSTM capacity limitation is in effect whenever there are more than $K$ elements in a display. To derive the probability of encoding a particular element of such a display, one must consider all combinations of elements filling up the VSTM, thus

$$\mathrm{p_E}(\, x \mid K < |S|,\, t > t_0\,) = v_x \sum_{j=0}^{K-1} \sum_{J \in P_j(\tilde{S})} \sum_{L \in P(J)} (-1)^{|L|} \frac{1-\exp(-[t-t_0]\nu)}{\nu} \tag{2}$$

where $\tilde{S} = S \setminus x$, $P(J)$ is the power set of $J$, $P_j(\tilde{S})$ is the set of subsets with cardinality $j$ of the power set of $\tilde{S}$, and $\nu = \sum_{m\in S} v_m - \sum_{l\in J} v_l + \sum_{k\in L} v_k$ , see Kyllingsbæk [5, 6], Dyrholm et al. [7] for derivation details and Dyrholm [8], Dyrholm et al. [7] for an algorithm that runs through the power set of a given elemental cardinality.

Only a limited amount of resources are available for processing a display, hence in multi-element displays the processing rate of an element depends on the other elements of the display. The processing rate of an element becomes a fraction of the total processing capacity according to its attentional weight $w_x$ relative to the sum of the attentional weights of all elements in the display

$$v_x = C \frac{w_x}{\sum_{z\in S} w_z} \tag{3}$$

where $C$ is the total processing capacity (elements per second). We use the software package of Dyrholm et al. [7] for estimating $K$, $t_0$, $C$, and $\{w_z\}$ from behavioral data consisting of Partial Report and Whole Report trials in which the task of the subject is to report target stimuli while ignoring potential distractor stimuli — the task is known as Whole Report when no distractors are displayed, but known as Partial Report when distractors are present, see also Shibuya and Bundesen [9].

## SINGLE-TRIAL LATENT VARIABLE INFERENCE ON A PARTIAL REPORT TRIAL

We now turn to the novel contribution of this paper: Using the observed response $R$ (a set of reported target stimuli) from a *single* Partial Report trial, we retrodict latent variables pertaining to the attentional state of the subject on that particular trial. We address two questions about the specific trial: 1) How many distractors were encoded? 2) Which distractors were encoded?

### *Question 1: How many distractors were encoded on the trial?*

First we compute the joint probability of encoding $d$ distractors *and* reporting the set $R$ of targets *given* the VSTM capacity $K$

$$\mathrm{p}(d,R|K) = \begin{cases} 0 & , \ d+|R| > K \\ \mathrm{p}_1 & , \ d+|R| < K \\ \mathrm{p}_2 & , \ d+|R| = K \end{cases} \tag{4}$$

where

$$\mathrm{p}_1 = \sum_{J \in P_d(D)} \prod_{k \in J \cup R} [1-\exp(-v_k[t-t_0])] \prod_{l \in S \setminus (J \cup R)} \exp(-v_l[t-t_0]) \tag{5}$$

$$\mathrm{p}_2 = \sum_{J \in P_d(D)} \sum_{j \in J \cup R} v_j \sum_{L \in P((J \cup R) \setminus j)} (-1)^{nL} \frac{1-\exp(-v[t-t_0])}{v} \tag{6}$$

$$\tag{7}$$

where $v = v_j + \sum_{l \in S \setminus (J \cup R)} v_l + \sum_{k \in L} v_k$. The VSTM capacity pdf is estimated by the software of Dyrholm et al. [7] and we can then out-marginalize $K$

$$\mathrm{p}(d,R) = \mathrm{p}(d+|R| < K) \times \mathrm{p}_1 + \mathrm{p}(d+|R| = K) \times \mathrm{p}_2 \tag{8}$$

Then the pdf over the number of distractors encoded given the report is found by normalizing

$$\mathrm{p}(d|R) = \frac{\mathrm{p}(d,R)}{\sum_d p(d,R)} \tag{9}$$

hence the number of distractors encoded on the trial is expected to have been $E[d|R] = \sum_d d\, \mathrm{p}(d|R)$.

*Question 2: Was a particular distractor encoded on the trial?*

First we derive the joint probability that distractor $i$ is encoded *and* $d$ distractors are encoded *and* the response $R$ is observed *given* $K$

$$\mathrm{p}(i,d,R|K) = \begin{cases} \mathrm{p}_0 & , |R|+d<K \\ \mathrm{p}_1 & , |R|+d=K \end{cases} \tag{10}$$

where $\mathrm{p}_0$ denotes the situation that the VSTM is not filled up

$$\mathrm{p}_0 = \sum_{J\in P_{d-1}(D\setminus i)} \prod_{j\in(R\cup J\cup i)} [1-\exp(-v_j[t-t_0])] \prod_{l\in S\setminus(R\cup J\cup i)} \exp(-v_l[t-t_0]) \tag{11}$$

and $\mathrm{p}_1$ denotes the situation that VSTM is filled up

$$\mathrm{p}_1 = \sum_{J\in P_{d-1}(D\setminus i)} \sum_{j\in J\cup R\cup i} v_j \sum_{L\in P([J\cup R\cup i]\setminus j)} (-1)^{nL} \frac{1-\exp(-v[t-t_0])}{v} \tag{12}$$

where $v = v_j + \sum_{l\in S\setminus(J\cup R\cup i)} v_l + \sum_{k\in L} v_j$. We can then derive the probability that the distractor was encoded *given* the actual response from the trial

$$\mathrm{p}(i|R) = \frac{1}{\mathrm{p}(R)} \sum_{d=1}^{|D|} \mathrm{p}_0 \times \mathrm{p}(K>d+|R|) + \mathrm{p}_1 \times \mathrm{p}(K=d+|R|) \tag{13}$$

where the VSTM capacity pdf is estimated by the software of Dyrholm et al. [7].

## EXAMPLE: COMBINING SINGLE-TRIAL BEHAVIORAL MODELING WITH EEG

### Behavioral paradigm

The CombiTVA paradigm of Vangkilde et al. [10, 11] employed here was a combination of two classic paradigms used to estimate visual attention capacity (the whole report paradigm; see Sperling [12]) and visual selectivity (the partial report paradigm; see Shibuya and Bundesen [9]), respectively. In the CombiTVA all trials (N=324) followed the same basic outline: every trial was initiated by a red fixation cross at the centre of the screen. After 1000 ms a letter display with six possible stimulus locations on an imaginary circle (r=7.4 visual degrees) around the fixation cross was flashed briefly. Displays consisted of either six or two red letters (whole report trials) or two red and four blue letters (partial report trials, N=81). Six-letter displays were presented for one of six exposure durations (10, 20, 50, 80, 140, or 200 ms), whereas all other displays were presented for 80 ms. Letter displays were terminated by a pattern mask covering all possible stimulus locations to effectively stop processing of the letters. All trial types were intermixed, and the task of the participant was to report as many red target letters as possible while ignoring the blue distractor letters. The letters for each display were chosen randomly without replacement from a set of 20 capital letters [ABDEFGHJKLMNOPRSTVXZ] and presented in the font Ariel broad (size: 68 pixels).

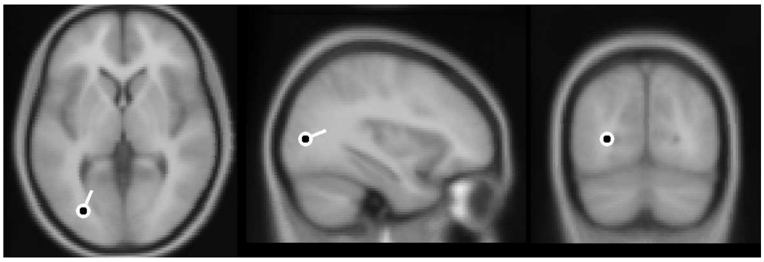

**FIGURE 1.** Current dipole localized to the Left Middle Occipital Gyrus. The directional icon is laid over a standard volume.

## Functional imaging with EEG

EEG was recorded from a female subject (age 25) using a 256 electrode BioSemi ActiveII system at a sampling rate of 2048 samples per second (bandwidth=$2048/5$ Hz) inside a shielded room. The EEG was high-pass filtered at 1 Hz, re-referenced to the average electrode, then down-sampled (after anti-aliasing low-pass filtering) to a sampling rate of 256 samples per second.

The data was then decomposed into independent components using the Infomax ICA algorithm implementation that comes with the EEGLAB software package [13, 14, 15]. A strong ICA component (ranked third in terms of power) with a scalp topography focus over the left occipital cortex was selected for further analysis: an area which has previously been proposed to reflect discriminative processing [16]. The power spectral density of the component indicated oscillatory activity with a peak around 12 Hz.

A single dipole was fitted to the component topography using the DIPFIT[1] plug-in for EEGLAB (based on principles of [17]). The dipole was localized to Talairach coordinates $(x, y, z) = (-29, -76, 10)$ which, by the aid of [18], can be characterized as the Left Middle Occipital Gyrus (see for example [19]). FIGURE 1 shows an iconic visualization of the dipole over a standard brain. The EEG/ICA component showed event-related attenuation of 12 Hz activity following the onset of the mask. In the following, we adopt the notion of Pfurtscheller [20] who used the term "event-related desynchronization" to characterize such observations.

## Results

To test the hypothesis that the observed event-related desynchronization of the occipital EEG/ICA component was partially due to processing load modulation, we used a Generalized Linear Model (GLM) with the response variable being 12 Hz power (analysis band: 10–14 Hz) in the Partial Report trials (N=81) following mask-onset (latency window: 0–250 ms). The power of the mentioned time-frequency window was quanti-

---

[1] DIPFIT kindly provided by Robert Oostenveld of Donders Institute for Brain, Cognition and Behavior, Center for Cognitive Neuroimaging.

**TABLE 1.** Likelihood ratio test versus the null hypothesis that no explanatory variables were useful in predicting the power of 12 Hz oscillations in the Left Middle Occipital Gyrus.

| **Distractors** | **Targets** | $-2\log\Lambda$ | $\Pr(>\chi^2)$ |
|---|---|---|---|
| Predicted | Observed | 5.6 | .06 |
| Predicted | Predicted | 5.0 | .08 |
| Inferred | Observed | 5.8 | .06 |
| Inferred | Predicted | 7.9 | .02* |

* Significant at 5% level. Not corrected for multiple comparisons.

fied by using multi-tapered spectral analysis with two Slepian sequences as defined by the 250 ms temporal window with a frequency analysis bandwidth of 4 Hz; see Slepian [21] for a definition of Slepian sequences. We assumed Gaussian noise, and a logarithmic link function so that the 12 Hz log-amplitude (which is proportional to the logarithm of the power) was assumed linear in the GLM parameters. We had two explanatory variables representing the targets: 1) Observed, i.e. number of targets reported by the subject, and 2) Predicted, i.e. model fit. We also had two variables representing the distractors: 1) Predicted, i.e. model fit, and 2) Inferred, i.e. computed as per Question 1. TABLE 1 summarizes the result of testing each combination of explanatory variables against the null hypothesis $H_0$ that no explanatory variables were useful. The use of Inferred Distractors yielded a more extreme statistic compared to using Predicted Distractors. The only significant combination was Inferred Distractors with Predicted Targets. This model accounted for 13% of the response variable's variance. Both model coefficients were negative indicating that increased processing load, targets and distractors alike, had a damping effect on the oscillations. This result supports our claim that the inferenced varibles can be useful in an account for single-trial cognitive fluctuations.

## CONCLUSION

In this paper we showed how latent variables can be inferred on a single-trial basis using TVA-based behavioral modeling. Further, we presented an example of how these inferred variables explained trial-by-trial fluctuations in an EEG recording. Specifically, we found that the event-related desynchronization of an EEG/ICA component localized to the Left Middle Occipital Gyrus was proportional to the inferred number of distractors reaching the VSTM on a given Partial Report trial.

## REFERENCES


1. M. G. Philiastides, and P. Sajda, *J. Neurosci.* **27**, 13082–13091 (2007).
2. M. Mørup, K. H. Madsen, and L. K. Hansen, Latent causal modelling of neuroimaging data, Tech. rep., Technical University of Denmark (2009).

3. G. J. Siegle, S. R. Steinhauer, V. A. Stenger, R. Konecky, and C. S. Carter, *Neuroimage* **20**, 114–124 (2003).
4. C. Bundesen, *Psychological Review* **97**, 523–47 (1990).
5. S. Kyllingsbæk, Item based fitting of whole report data, Tech. rep., University of Copenhagen (2001).
6. S. Kyllingsbæk, *Behavior Research Methods* **38**, 123–133 (2006).
7. M. Dyrholm, S. Kyllingsbæk, T. Espeseth, and B. Bundesen, *Submitted* (2011).
8. M. Dyrholm, A recursive algorithm for forming the constrained elemental cardinality power set, Tech. rep., University of Copenhagen (2009).
9. H. Shibuya, and C. Bundesen, *Journal of Experimental Psychology: Human Perception and Performance* **14**, 591–600 (1988).
10. S. Vangkilde, C. Bundesen, and J. Coull, "Differential effects of nicotine on discrete components of visual attention," in *Proceedings of the International Neuropsychological Society Mid-year Meeting*, 2009, p. 53.
11. S. Vangkilde, C. Bundesen, and J. T. Coull, *In preparation* (2011).
12. G. Sperling, *Psychological Monographs* **74** (1960).
13. S. Makeig, A. Bell, T.-P. Jung, and T. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, edited by D. Touretzky, M. Mozer, and M. Hasselmo, MIT Press, 1996, vol. 8, pp. 145–151.
14. T.-W. Lee, M. Girolami, and T. J. Sejnowski, *Neural Comput.* **11**, 417–441 (1999).
15. A. Delorme, and S. Makeig, *Journal of neuroscience methods* **134**, 9–21 (2004).
16. S. J. Luck, *An Introduction to the Event-Related Potential Technique*, The MIT Press, Cambridge, MA, 2005.
17. M. Scherg, "Fundamentals of dipole source potential analysis," in *Advances in Audiology*, edited by F. Grandori, M. Hoke, and G. Romani, S. Karger, Basel, 1990, vol. 6, pp. 40–69.
18. F. A. Nielsen, and L. K. Hansen, *Artif Intell Med* **30**, 141–151 (2004).
19. J. Downar, A. P. Crawley, D. J. Mikulis, and K. D. Davis, *Nat. Neurosci.* **3**, 277–283 (2000).
20. G. Pfurtscheller, *Electroencephalogr Clin Neurophysiol* **43**, 757–760 (1977).
21. D. Slepian, *Bell System Tech. J.* **57**, 1371–1429 (1978).