Revisiting the orthographic prediction error for a better understanding of efficient visual word recognition

Wanlu Fu (wanlu.fu@uni-koeln.de) & Benjamin Gagl (benjamin.gagl@uni-koeln.de)
Self Learning Systems Lab at the Department of Special Education and Rehabilitation
Frangenheimstraße 4, Cologne, Germany

Abstract

Recent evidence suggests that readers optimize low-level visual information following the principles of predictive coding. Based on a transparent neurocognitive model, we postulated that redundant visual signals are removed, allowing readers to focus on the informative aspects of the visual percept, i.e., the orthographic prediction error (oPE). Here we test alternative oPE implementations by assuming all-or-nothing signaling units based on multiple thresholds (i.e., output modality of a neuron). Further, we tested if predictions are signaled from one or multiple units. For model evaluation, we compared the new oPEs with each other and against the original formulation based on behavioral and electrophysiological data (EEG at 230, 430 ms). We found the highest model fit for the oPE with a 50% threshold integrating multiple prediction units for behavior and the late EEG data. The early EEG data was still explained best by the original hypothesis. Thus, the new formulation is adequate for late but not early neuronal signals indicating that the representation, which likely implements lexical access, changes over time.

Keywords: Visual word recognition; Visual-orthographic processing; Predictive coding; EEG; Behavior

Introduction

Efficient readers read up to 300 words per minute (Brysbaert, 2019) by picking up visual information every 200-250 ms (Gagl, Gregorova, et al., 2022; Siegelman et al., 2022). To achieve fast access to word meanings, readers implement efficient representations (Gagl et al., 2020) and integrate words in their sentence/text context to generate expectations for upcoming words (Hawelka, Schuster, Gagl, & Hutzler, 2015; Heilbron, Richter, Ekman, Hagoort, & de Lange, 2020; Hofmann, Remus, Biemann, Radach, & Kuchinke, 2022).

Here we revisit our formulation of efficient representations of visual information based on knowledge-based orthographic expectations (Gagl et al., 2020) following the principles of predictive coding (Rao & Ballard, 1999). In this implementation, we assumed signaling of graded values on the prediction and prediction error level. Here, we tested an alternative implementation considering a previously neglected constraint, i.e., neurons implement a binary all-or-nothing signal and cannot implement a graded signal (Qin et al., 2020; Saszik & DeVries, 2012). We used thresholds (from 10-90% in 10% steps) to implement the adaptation (see examples in Fig. 1).

With this change, new hypotheses concerning the structure of a potential neuronal circuit evolve. We here implement simplified assumptions compared to previous models (Bastos et al., 2012). We assume that the bottom-up input to a prediction error unit is the information of one pixel from a word image (Fig. 1). In line with Gagl et al.(2020), readers inhibit the sensory input when they predict the sensory input based on orthographic knowledge. Here, a question arises from assuming all-or-nothing units: Do the *prediction error* units integrate signals from one or multiple *prediction* units? We implement multiple units (i.e., a graded prediction based on a mean; when one of multiple prediction units fires, the prediction is weaker cp. to when all fire) to test if one integrates prediction signals across units (Fig. 1A). In contrast, assuming one prediction unit, we implement a binary prediction (Fig. 1B) as then the prediction would be there or not. Here, we contrast the original model with the two structural variants of the new implementation for each threshold based on three datasets (lexical decision behavior, EEG at posterior sensors around 230 ms, and frontal sensors around 430 ms).

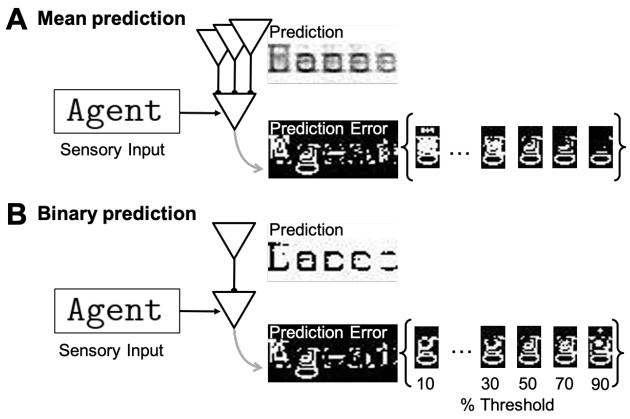

Figure 1: Model overview. (A) Structure assuming multiple prediction units (i.e., mean prediction) and (B) a single prediction unit (i.e., binary prediction) with orthographic prediction error examples for “g” from multiple thresholds.

Method

We adopt the model comparison analysis procedure and three datasets from Gagl et al. (2020) to evaluate the new orthographic prediction error (oPE) variants (behavioral lexical decision data; EEG from posterior electrodes 230 ms and frontal electrodes 430 ms after stimulus onset; find here: https://osf.io/d8yjc/). All stimuli had five letters (Behaviour: 800 Words & non-words; EEG: 200 Words & non-words),

and participants were typically reading native speakers (Behaviour: N = 35; EEG: N = 31).

To identify the best model for the three datasets, we implement both variants from Figure 1 using thresholds from 10-90% of the signal in 10% steps (see examples in Fig. 1). In other words, a prediction error for a pixel was only generated when the difference between the sensory input and the prediction exceeded the threshold (e.g., if the sensory input is 1 and the prediction is .2, the prediction error will be 1 as the 50% threshold is lower at .5 than the difference of the prediction and the sensory input with .8; if the prediction, in this case, is .6 the prediction error will be 0). We summed gray values from the prediction error images to obtain a numeric predictor for each stimulus. With this measure, we estimated linear mixed models (Bates, Mächler, Bolker, & Walker, 2014), including the oPEs as predictors to describe response times and EEG amplitudes. Also, the previous analysis (see Gagl et al. 2020) suggested estimating the interaction of the oPE and word lexicality (i.e., word or non-word) for the later time window and the behavioral data (additional parameters for EEG data: lexicality, & number of pixels, i.e., amount of the sensory input; for the behavioral data: word/non-word, number of pixels, word frequency, and decision accuracy). For the model fit comparison, we used the Akaike Information Criterion (AIC) in relation to the null model without an oPE (Akaike, 1973).

Results

Correlations between the new and the original oPEs in both stimulus sets showed low similarity for the low percentage thresholds (10-20% thresholds, range: r = .07 to .32) and high similarity for the other implementations (30-90% thresholds range: r = .52 to .92).

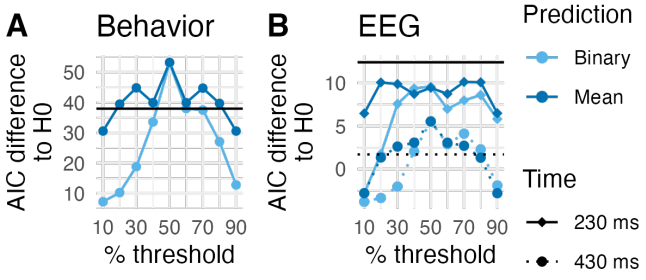

Figure 2: Model comparisons results based on (A) lexical decision response times for the original oPE (solid black line) and new implementations, including all thresholds and both prediction assumptions. All AICs presented here are against a version of the linear mixed model without an oPE predictor (Null hypothesis, H0). (B) Same but for EEG amplitudes at 230 ms and 430 ms. The AIC difference for the original oPE is presented as solid (230 ms) and dotted line (430 ms).

Behavioral results. Behavioral evaluations indicated the highest model fit at a threshold of 50% for the Mean prediction with only one AIC point difference to the Binary prediction (see Fig. 2A). Generally, higher model fits are found for the Mean prediction. Also, compared to the original oPE, all but two oPEs (10 & 90% thresholds; 2 of 9) based on the Mean prediction and all but seven (lower AICs for 7 of 9) based on the Binary prediction showed higher model fits.

Electrophysiological results. The original oPE had the highest model fit at amplitudes at 230 ms. Still, all but one new oPE formulation (10% threshold, binary prediction) increased the model fit significantly in contrast to the null model (i.e., AIC difference $>$ 3; see Fig. 2B). Again, oPEs based on the Mean prediction had higher model fits in all but one case (40%).

At 430 ms, the model comparison pattern was highly similar to the behavioral results (cp. Binary: r = .97; Mean: r = .93 vs. Binary: r = .76; Mean: r = .75 for the 230 ms time window). Although the general AIC differences have been smaller than in the early time window, we found a clear peak for the 50% threshold oPE of both prediction assumptions that was higher cp. to the original oPE. Again, we found higher AIC differences for oPEs based on the Mean prediction assumptions (6 of 9).

Discussion

The exploration of alternative oPE formulations, increasing neuronal plausibility (i.e., all-or-nothing binary oPE), and investigating simple neuronal architectures (mean and binary predictions), found that an all-or-nothing binary oPE implemented with a 50% threshold and a prediction that integrates multiple units (Mean prediction) best explained behavioral performance and brain potentials at 430 ms (Frontal sensors). Correlations of the model comparison results from the very different data (i.e., different measures, participants, and stimuli) indicated high similarity. Comparing Mean and Binary prediction assumptions, we found high similarity and only small model fit differences for the best fitting model (at 50% threshold). Still, for most other thresholds, mean predictions resulted in higher model fits. In contrast, the original oPE implementation explained the early brain responses best. This difference (i.e., early vs. late) could indicate a transformation of the oPE representation over time.

The critical difference between the behavior, late and early brain potentials is that only for the earlier time window, readers did not achieve lexical access yet. The 50% threshold oPE, thus, likely represents the final form of the oPE representation that might allow accessing a lexical unit. To get a clearer picture of the change from the original to the 50% threshold oPE, new analyses need to investigate the entire EEG time course, including potential other processes involved before lexical access (Gagl, Weyers, & Mueller, 2021; Gagl, Richlan, et al., 2022; Lerma-Usabiaga, Carreiras, & Paz-Alonso, 2018; White, Palmer, Boynton, & Yeatman, 2019). A possible mechanism to explain the oPE change one needs to identify an architecture that can remove weak prediction errors of the graded original oPE and amplify prediction errors at the 50% threshold. Thus, readers likely use binary oPE representations to access word meanings that originate from a graded oPE representation.

References

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265.

Bastos, A., Usrey, W., Adams, R., Mangun, G., Fries, P., & Friston, K. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695-711. doi: https://doi.org/10.1016/j.neuron.2012.10.038

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4.* arXiv. doi: 10.48550/ARXIV.1406.5823

Brysbaert, M. (2019). How many words do we read per minute? a review and meta-analysis of reading rate. *Journal of memory and language*, *109*, 104047.

Gagl, B., Gregorova, K., Golch, J., Hawelka, S., Sassenhagen, J., Tavano, A., ... Fiebach, C. J. (2022). Eye movements during text reading align with the rate of speech production. *Nature human behaviour*, *6*(3), 429–442.

Gagl, B., Richlan, F., Ludersdorfer, P., Sassenhagen, J., Eisenhauer, S., Gregorova, K., & Fiebach, C. J. (2022). The lexical categorization model: A computational model of left ventral occipito-temporal cortex activation in visual word recognition. *Plos Computational Biology*, *18*(6), e1009995.

Gagl, B., Sassenhagen, J., Haan, S., Gregorova, K., Richlan, F., & Fiebach, C. J. (2020). An orthographic prediction error as the basis for efficient visual word recognition. *NeuroImage*, *214*, 116727. doi: https://doi.org/10.1016/j.neuroimage.2020.116727

Gagl, B., Weyers, I., & Mueller, J. L. (2021). Speechless reader model: A neurocognitive model for human reading reveals cognitive underpinnings of baboon lexical decision behavior. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2015). On forward inferences of fast and slow readers. an eye movement study. *Scientific reports*, *5*(1), 1–8.

Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & de Lange, F. P. (2020, jan). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, *11*(1). doi: 10.1038/s41467-019-13996-4

Hofmann, M. J., Remus, S., Biemann, C., Radach, R., & Kuchinke, L. (2022). Language models explain word reading times better than empirical predictability. *Frontiers in Artificial Intelligence*, *4*, 214.

Lerma-Usabiaga, G., Carreiras, M., & Paz-Alonso, P. M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the National Academy of Sciences*, *115*(42), E9981–E9990.

Qin, H., Gong, R., Liu, X., Bai, X., Song, J., & Sebe, N. (2020). Binary neural networks: A survey. *Pattern Recognition*, *105*, 107281.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79–87.

Saszik, S., & DeVries, S. H. (2012). A mammalian retinal bipolar cell uses both graded changes in membrane voltage and all-or-nothing na+ spikes to encode light. *Journal of Neuroscience*, *32*(1), 297–307.

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., ... others (2022). Expanding horizons of cross-linguistic research on reading: The multilingual eye-movement corpus (meco). *Behavior research methods*, 1–21.

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences*, *116*(20), 10087–10096.