Chapter 2

Fast Visual Specialization for Print and Reading Fluency
Chapter 2  Fast Visual Specialization for Print and Reading Fluency

Chapter 2

Fast Visual Specialization for Print and Reading Fluency: Using Generalized Additive Modelling to Study a Neurophysiological Marker of Emerging Dyslexia

Abstract | Background: When presented with print (words or word-like letter strings), fluent adult readers activate fast, specialized visual brain processes, which underlie the basis for efficient recognition of print input. At the neurophysiological level, specialized visual processing of print is indexed by the N170 component of the event-related potential (ERP): in alphabetic languages, print consistently elicits larger N170 responses than visual baseline (e.g., symbols), especially in the left hemisphere. Such a tuning effect develops rapidly within the first two years of schooling in typically developing children, and is believed to reflect reading expertise associated with one’s own language. The finding of absent/attenuated tuning effect in children with dyslexia, on the other hand, has generated expectations that the print N170 may help distinguish poor from normal readers at an early age.

Aims: In the current study, we thus investigated whether the N170 print-tuning effect has already emerged in beginning readers of Dutch (second graders), and, if yes, whether the size and the lateralization pattern of this tuning effect are systematically related to reading fluency.

Methods & Procedures: To fulfil these objectives, we recorded the ERPs of a group of Dutch children with varying degrees of reading fluency while they were performing a repetition detection task with words, pseudo words and symbol strings. In offline analysis of the ERP data, we used generalized additive modelling (GAM) to investigate the relation between reading score and the size, as well as the lateralization pattern of the N170 print-tuning effect. Such a novel statistical tool allows us to assess the complete, nonlinear shape of the ERP signal over time while controlling for systematic variations associated with subjects and items.

Results & Conclusions: The results of the GAM analyses demonstrated a robust N170 print-tuning effect (i.e., enhanced N170 for words compared to symbols) in the left hemisphere, as well as a positive, almost linear relation between reading ability and the size of the tuning effect. Overall, our findings lend support to the validity of the N170 print-tuning effect as a neurophysiological marker of emerging dyslexia.

Key words: visual word recognition, developmental dyslexia, N170 print-tuning effect

---

2This chapter was adapted from: Qin, R., Wieling, M., van der Spek, J., Maurits, N. M., & Maassen, B. (2015). Fast visual specialization for print and reading fluency: Using generalized additive modelling to study a neurophysiological marker of emerging dyslexia. Manuscript submitted for publication.
Chapter 2  Fast Visual Specialization for Print and Reading Fluency

2.1. Introduction

Developmental dyslexia is a neurobiologically-based learning disorder that impairs a child’s ability to read and write. Over the past few decades, the flourishing of neuroimaging techniques has significantly advanced our understanding of the neural substrates of reading and reading disorders. Such knowledge is key to identifying early precursors of dyslexia, and may eventually give rise to highly-targeted intervention programs (Posner & Rothbart, 2005).

One of the most intensively investigated parts of the reading network is the left ventral occipito-temporal cortex. This particular patch of the visual cortex is believed to underlie fast, automatized recognition of print (visually presented words or word-like letter strings), and is hence termed as the visual word form system (VWFS, Cohen et al., 2000). The VWFS appears to be tuned to print processing: in response to print, fluent adult readers activate fast, specialized visual brain processes in the VWFS, enabling efficient recognition and categorization of the stimuli (Brem et al., 2006, 2010; Dehaene, Cohen, Sigman, & Vinckier, 2005; van der Mark et al., 2009; Vinckier et al., 2007). Such print-specific activation, however, seems to be reduced in dyslexic adults and children (Helenius, Tarkiainen, Cornelissen, Hansen, & Salmelin, 1999; Kronschnabel et al., 2013; Paulesu, et al., 2001; Richlan et al., 2011; Rumsey et al., 1997; Shaywitz et al., 2002; van der Mark et al., 2009; Wimmer et al., 2010).

An apt tool for studying such transient visual-specialization processes is the event-related potential (ERP), which records electrical brain activity on a time scale of milliseconds. Previous research has pointed to the N170 component of the ERP as the electrophysiological correlate of print tuning, which is an occipito-temporal negativity peaking between 140 and 220 ms after stimulus onset. The N170 has long been known to respond more strongly to particular classes of stimuli than to others, depending on the visual expertise of the viewer. For example, bird experts show larger N170 to bird than to dog stimuli, whereas dog experts demonstrate the opposite pattern (Tanaka & Curran, 2001). Furthermore, familiarity with human faces results in enlarged N170 responses for face stimuli (Gauthier et al., 2003; Latinus & Taylor, 2006). Analogously, exposure to written language gives rise to N170 tuning effect for print: in fluent adult readers, the N170 is consistently found to be larger for print than for visually-matched control stimuli (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Brandeis, Lehmann, Michel, & Mingronem, 1995; Brem et al., 2005; Maurer et al., 2005a, 2005b, 2008; Rossion et al., 2003). In contrast to the object and the face N170, which typically show a bilateral pattern, the print N170 is lateralized to the left hemisphere (Brandeis et al., 1995; Brem et al., 2005; Rossion et al., 2003). Source localization studies have identified the left VWFS as the neuroanatomical correlate of the print N170 (Helenius et al., 1999; Proverbio & Zani, 2003).

Specifically, two levels of N170 tuning have been reported. A coarse form of tuning is reflected by larger N170 amplitudes for print stimuli (e.g., words or pseudowords) than for nonsense character strings made from symbols or icons. Such enhanced responses can be readily elicited using implicit reading tasks, indicating that the prelexical sensitivity of N170 is largely automatic (Bentin et al., 1999; Schendan, Ganis, & Kutas, 1998). A second, fine form of N170 tuning concerns differential processing of familiar and unfamiliar word forms (e.g., words vs. pseudowords, or high-frequency vs. low-frequency words): the former are easier to process and therefore evoke a smaller N170 than the latter (Curran, Tanaka, & Weiskopf, 2002; Hauk & Pulvermüller, 2004; Sereno, Brewer, & O’Donnell, 2003). Compared with the
coarse tuning, the fine tuning is less robust. The subtle differences in N170 amplitude between different subtypes of print stimuli can be observed in explicit reading tasks, such as lexical or semantic tasks (Bentin et al., 1999; McCandliss, Posner, & Givon, 1997), but not in implicit reading tasks (Bentin et al., 1999; Maurer et al., 2005 a, 2005b; Schendan et al., 1998; Wydell, Vuorinen, Helenius, & Salmelin, 2003).

The two forms of N170 tuning also differ in their developmental trajectories. While the fine tuning doesn’t fully develop until reading has become automatic (McCandliss & Noble, 2003), the coarse tuning emerges rather early, with its size peaking within the first two years of reading acquisition. In a longitudinal study by Maurer and colleagues (Maurer et al., 2006), preliterate kindergarten children showed identical N170 responses to words and symbol strings; after two years of reading education, however, enhanced N170 for words was reliably found in the same children. A comparison between the second graders and adults revealed that the coarse tuning effect attenuates during adulthood, presumably because reading has become increasingly proficient through intensive training. Taken together, these findings suggest that the N170 tuning effect for print, especially its coarse form, is a valid marker for emerging reading expertise. Therefore it is of particular interest to investigate N170 tuning in novice readers with dyslexia, as it may prove to be a valid neurophysiological indicator of the disorder. Supporting this view, evidence has been accumulating that the left-lateralized print N170 develops rapidly with learning to read in typically developing children, but not in dyslexic children (Araújo et al., 2012; Eberhard-Moscicka et al., 2015; Maurer et al., 2007, 2011).

In the current study, we aim to investigate the relation between N170 print-tuning effect and reading fluency in novice readers of Dutch. Our prediction is that children with better reading performances show larger print-tuning effects, especially in the left hemisphere. A particular difficulty in ERP studies with children is to define the time window wherein peak or mean amplitude of the component under investigation is to be measured. This is because ERP signal in children tends to follow a different time course from that in adults, and is characterized by greater individual variations. In this study we use generalized additive modelling (GAM; Wood, 2006) to analyse the N170 tuning effect for print, a statistical tool for modelling nonlinear relations. With GAM it is possible to assess the entire time-course of the ERP signal (which exhibits a nonlinear shape). There is hence no need for a pre-defined time window of analysis. As a regression method, GAM is also capable of accommodating numeric predictors. Such a characteristic allows us to use raw reading scores to index reading ability, rather than resorting to group dichotomization (poor-reader vs. control), which often proves to be arbitrary when participants are yet too young for diagnosis. Since GAM is relatively new in ERP research, we provide a brief introduction to this statistical tool in the Methods section.

2.2. Methods

2.2.1. Participants

Twenty-one children (9 males and 12 females; mean age [±SD] = 7.7 [±0.46] years) participated in the current study. All children were recruited from regular primary schools, and were tested during the second half of Grade 2. Teachers were asked to refer especially poor readers for this study. Informed consent was obtained from both parents prior to inclusion. Parents also filled out questionnaires regarding neurological disorders, handedness and visual acuity. All participants were neurologically healthy,
right-handed, and had normal or corrected-to-normal vision. The study was approved by the Medical Ethical Review Board of the University Medical Center Groningen (ABR nr NL43354.042.13).

Before participating in the ERP experiment, the children were tested (on a separate day) for reading fluency and reading-related cognitive abilities. A comprehensive behavioral test battery was administered, which comprised assessments of word reading (Drie-Minuten Toets, hereinafter referred to as DMT; Verhoeven, 1995), phoneme deletion, rapid automatized naming (RAN test; Denckla & Rudel, 1976), and orthographic knowledge. For statistical analyses of the ERP data, scores on the DMT test were used to index reading ability, which is a standardized word-reading test in Dutch. All children had normal nonverbal intelligence based on performance on the Matrix Reasoning subtest of the Wechsler Preschool and Primary Scale of Intelligence test (WPPSI; Wechsler & Naglieri, 2008).

### 2.2.2. Materials

The participants performed a one-back repetition detection task with three conditions: Dutch words, pseudowords and symbols. The Dutch words were high-frequency nouns (average occurrences per million: 2.75; Corpus Gesproken Nederlands [Spoken Dutch Corpus]). All words were selected from Grade-one textbooks to ensure the children’s familiarity with them. The pseudowords were pronounceable pseudowords. The symbol strings consisted of simple geometric symbols, similar to those used in previous studies (Brem et al. 2005; Maurer et al., 2006, 2007, 2008). The stimuli were matched for string length (mean length in cm [±SD]: word, 12.4 [±2.1]; pseudoword: 11.8 [±2.3]; symbol, 13.1 [±1.9]) and the number of characters in a string (3 to 5) across conditions.

![Stimuli and experimental procedure](image)

**Fig. 2.1: Stimuli and experimental procedure.** Stimulus duration: 700 ms; ISI: 1000 ms.

The stimuli were shown in black on a white background in the centre of the screen. For each condition, there were two blocks of 48 items; for each block, 8 immediate repetitions were pseudo-randomly presented as targets (17%), avoiding two
immediate repetitions in a row. The items were shown for 700 ms, followed by an inter stimulus interval (ISI) of 1000 ms. Participants were seated 1.2 meters away from the computer screen, and were instructed to press a joystick button as quickly as possible after immediate stimulus repetition (see Fig. 2.1). The N170 is an automatic response to visual stimulation that can be elicited irrespective of the subjects’ attention or behavioural task. The purpose of having a task was to ensure that the participants viewed the stimuli. Overall, the children were able to detect immediate repetitions with fair accuracy across all conditions (average accuracy [%±SD]: word, 89.7 [%±7.8]; pseudoword: 84.3 [%±9.1]; symbol: 81.6 [%±11.7]), indicating that they were attending to the stimuli during the experiment.

2.2.3. ERP recording and analysis

The electroencephalogram (EEG) was recorded on an ASA-Lab system, using a 64-channel WaveGuard cap with sintered Ag/AgCl electrodes (ANT Neuro Inc., Enschede, the Netherlands). Additional electrodes were used to record horizontal (HEOG; electrodes positioned at the outer canthus of each eye) and vertical eye movements (VEOG; electrodes positioned above and below the left eye). Impedance was kept below 20kΩ. The EEG was digitized online at a sampling rate of 512 Hz and was referenced to the average of all channels.

For offline analyses, the acquired EEG data were corrected for eye blinks and horizontal eye movements using principal component analysis (Ille, Berg, & Scherg, 2002). The corrected data were subsequently band-pass filtered (0.5 to 40 Hz) and segmented into trials of 1200 ms duration, including a 200 ms baseline. Trials with extreme amplitudes (larger than $\pm 100 \mu V$) were rejected. The remaining trials were baseline corrected and down-sampled to 100 Hz to speed up further analyses. The data were then exported from ASA-Lab to the statistical package R, including only non-target trials (target trials were excluded because they were likely to be contaminated by finger movements involved in button pressing). Specifically, three homologue pairs of occipito-temporal channels, i.e., P7/P8, O1/O2, PO7/PO8, were selected for statistical analyses according to the maximum of the potential distribution on topographic maps. Mean amplitude for each hemisphere was computed by averaging across selected channels per hemisphere. As shown in Fig. 2.2, grand average ERPs at these occipito-temporal channels reveal a clear negativity between approximately 175 and 275 ms in all conditions. This negativity is more prominent in the word and the pseudoword condition than in the symbol condition, whereas virtually no difference can be observed between the word and the pseudoword condition. For the analysis we hence focused on the crucial contrast between word and symbol, which indexes the coarse print-tuning effect.

2.2.4. Statistical analysis

2.2.4.1. Generalized additive modelling

We used Generalized Additive Modelling (GAM; Wood, 2006) to analyse the exported ERP data. In contrast to linear regression, which assumes linearity in the relation between individual predictors and the dependent variable, GAM provides flexible tools for modelling nonlinear relations. An important advantage of GAM is that the form of the nonlinearity does not need to be specified in advance (such as $y = x^2$, defining a parabola). Rather, GAM determines the optimal shape of the
nonlinearity automatically, in such a way that it strikes a balance between under- and over-smoothing (via generalized cross-validation or relativized maximum likelihood; see Wood, 2006 for more details).

In GAM, via smooth functions termed “splines”, it is possible to model individual (nonlinear) predictors, as well as nonlinear interaction between multiple predictors. An apt tool to assess the nonlinear effect of a single predictor is the thin plate regression spline, which models the nonlinear pattern as a combination of several simple functions (a linear line, a quadratic function, a cubic polynomial, etc.). To link multiple predictors, a tensor product spline can be used, which, by combining cubic polynomials in all dimensions under investigation, generates a wiggly surface representing the nonlinear interaction between the predictors (see Wieling, Montemagni, Nerbonne, & Baayen, 2014 for an application of the tensor product spline in dialectology). Crucially, GAM always determines the amount of smoothing automatically; the user only needs to specify the upper bound on the complexity/degree of nonlinearity of the surface (Tremblay & Newman, 2015). In recent years, GAM has been applied successfully to many domains, including dialectology (Wieling et al., 2014; Wieling, Nerbonne, & Baayen, 2011) and psycho- and neurolinguistics (Kryuchkova, Tucker, Wurm, & Baayen, 2012; Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015; Tremblay & Baayen, 2010). A more detailed introduction to GAM can be found in Baayen, Kuperman, and Bertram (2010).

Fig. 2.2: Average ERPs at occipito-temporal channels.

2.2.4.2. Applying GAM to ERP data

GAM provides an innovative approach for analysing time-series data consisting of measurements made over consecutive time points, such as ERP data. In traditional ANOVA models, analysis of the ERP signal involves averaging over multiple trials
for each subject. Peak amplitude or mean amplitude in a given time window is subsequently measured on the averaged waveform. As a consequence, in such an approach only one data point per trial is considered and the temporal dynamics of the data are lost. The cause of this problem lies in the fact that the ANOVA approach (but also linear mixed-effects regression) cannot accommodate nonlinearities, such as the shape of the ERP signal over time. We therefore turned to GAM, which allows us to model these nonlinearities directly. As an example, in Fig. 2.3, a thin plate regression spline models the nonlinear relation between time and the amplitude of the ERP signal (measured in microvolt), yielding a nonlinear curve describing the pattern of the brain potentials during an interval of 400 ms after stimulus onset. Notably this curve closely resembles the grand average ERPs (Fig. 2.2), with a clear negativity peaking between 175 and 275 ms.

Another advantage of GAM over ANOVA is that there is no need to dichotomize quantitative variables that are inherently continuous. When participants are too young for diagnoses of dyslexia (i.e. approximately before grade two in the Netherlands), a common practice in line with the ANOVA approach is to group participants into “poor-readers” vs. “controls” on the basis of their performance on reading and/or other cognitive tests. This dichotomy, which in many cases is somewhat arbitrary, is likely to decrease statistical power and increase the likelihood of generating spurious significance (McCallum, Zhang, Preacher, & Rucker, 2002; Tremblay & Newman, 2015). Since GAM allows us to model nonlinear interactions among multiple continuous predictors (see above), we use raw reading scores instead of the grouping factor “poor-reader vs. control” to index the children’s reading ability. As an example, in Fig. 2.4, a tensor product spline models the nonlinear ERP pattern over time and across reading scores. The contour lines represent isolines connecting combinations of time and reading score modelled to have the same amplitude. The colours correspond...
to the isolines, with blue, green, yellow and light grey indexing increasing values. For example, the blue areas around 200 ms indicate negative responses across all reading scores, with deeper shades of blue indicating more negative values; between 300 and 400 ms, participants with low (30 – 50) or high (70 – 100) reading scores showed positive responses (coded with yellow or light grey), whereas those with moderate reading scores (50 – 70) showed responses close to zero microvolt (coded with green).

![Contour plot of ERP data](image)

**Fig. 2.4: Visualizing the nonlinear interaction between Time and Reading score (DMT).** The contour plot shows a regression surface of the amplitude of the ERP signal as a function of Time and Reading score, modeled with a tensor product spline. The contour lines represent isolines connecting combinations of Time and Reading score modeled to have the same amplitude. The colors correspond to the isolines, with blue, green, yellow and light grey indexing increasing values.

### 2.2.4.3. Modelling random effects in GAM

As a mixed-effect regression model, GAM can take into account random-effect factors. In regression analyses and other methodologies, fixed-effect factors refer to variables with a limited number of levels, all of which are included in the data (e.g., gender). On the other hand, random-effect factors are factorial predictors with a large number of levels, from which a random sample is drawn and investigated. In the current study, random-effect factors include subject and item (i.e., we could have included other children and a different set of words/pseudowords/symbols).

Since our subjects and items are randomly sampled from a larger population, they are likely to introduce systematic variation. In mixed-effects regression models, random intercepts and slopes are used to account for this variation and thereby prevent Type I errors in estimating the significance of the predictors (Baayen, Davidson & Bates, 2008; Jaeger, 2008). In GAM, factor smooths, a non-linear analogue to random intercepts/slopes, can be used to account for nonlinear individual variability and prevent overconfident (i.e., too thin) confidence bands (see Meulman et al., 2015). For instance, each child may show a specific (nonlinear) ERP pattern...
over time in each condition and hemisphere. The GAM approach takes this variability into account, thereby increasing the generalizability of the results.

### 2.3. Results

In the current study, we used GAM to investigate the relation between reading performance and the size, as well as the lateralization pattern of the N170 print-tuning effect. The analyses were performed using the mgcv package (version 1.8.4; Wood, 2006) in R (version 3.1.2). The minimally adequate model was determined by means of model comparison. Specifically, ML (maximum likelihood estimation) scores were used to compare models differing in fixed effects, and fREML (fast restricted maximum likelihood estimation) scores were used to compare models differing in random effects. Since GAM reports negative likelihood, a reduction in ML/fREML scores indicates an increase in the goodness of fit and the generalizability of the model (Wieling et al., 2011, 2014; Wood, 2006).

To start with, we fitted a baseline model with a nonlinear pattern over Time as the only fixed-effect factor, and Subject and Item as random-effect factors. In subsequent models, Reading score (DMT), Hemisphere and Condition were added step-by-step while the random-effect factors were held constant. The inclusion of each fixed-effect factor, either categorical or numeric, was supported by Chi-square tests of ML scores (all $p$s < 0.001), i.e., the extra complexity was warranted given the improvement in model performance. Similarly, to assess whether additional complexity was needed in the random-effect structure (e.g., to allow for hemispheric differences in individual ERP patterns over time), the fixed-effect factors were kept constant (i.e., those identified to be necessary in the previous step) and Chi-square tests of fREML scores were applied. The resulting model with all significant predictors and verified random-effect structure is specified as follows:

\[
\mu V = \text{tensor} (\text{Time, DMT, by} = \text{Hemisphere} \times \text{Condition}) \\
+ \text{Hemisphere} \times \text{Condition} \\
+ \text{factor smooth} (\text{Time, Subject} \times \text{Hemisphere}) \\
+ \text{factor smooth} (\text{Time, Item} \times \text{Hemisphere})
\]

Our dependent variable, i.e., the amplitude of the brain potential (in $\mu V$), is measured over the first 400 ms after stimulus onset. We include a nonlinear interaction between Time (in milliseconds) and Reading score (measured by the DMT test) via a tensor product spline, which models a similar surface as visualized in Fig. 2.4. The by-part in the tensor product spline allows the surface to vary for each of the four combinations of Hemisphere (left or right) and Condition (word or symbol). As the tensor product splines are approximately centred (i.e., on average equal to 0), overall differences (across the entire time window) between the four combinations are accommodated by the separate interaction between Hemisphere and Condition. Finally, the model includes factor smooths per subject and per item, which allow variability in the ERP signal over time associated with individual subjects and items (separately per hemisphere). This final model explains approximately 30% of the variance of our dependent variable. The significance and associated statistics of the fixed-effect predictors and the random-effect structure are summarized in Table 2.1 and Table 2.2, respectively. A detailed interpretation of the model estimates is provided below.
Interpreting fixed-effect predictors

Table 2.1: Fixed effects of the final model

| Parametric Coefficients       | Estimate | Std. error | t value | Pr(>|t|) |
|-------------------------------|----------|------------|---------|----------|
| Intercept (Left Symbol)       | 5.83     | 0.99       | 5.84    | 5.23e-09*** |
| Left Word (vs. Intercept)     | -2.02    | 0.44       | -4.60   | 4.16e-06*** |
| Right Symbol (vs. Intercept)  | -1.69    | 1.20       | -1.40   | 0.16     |
| Right Word (vs. Intercept)    | -3.15    | 1.18       | -2.66   | 0.00776**  |

Regression Surfaces

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>tensor (t, DMT): Left Symbol</td>
<td>9.14</td>
<td>10.60</td>
<td>8.82</td>
<td>3.31e -15***</td>
</tr>
<tr>
<td>tensor (t, DMT): Left Word</td>
<td>8.08</td>
<td>8.27</td>
<td>11.45</td>
<td>6.2e -14***</td>
</tr>
<tr>
<td>tensor (t, DMT): Right Symbol</td>
<td>12.34</td>
<td>14.87</td>
<td>7.05</td>
<td>1.6e -15***</td>
</tr>
<tr>
<td>tensor (t, DMT): Right Word</td>
<td>7.82</td>
<td>8.16</td>
<td>10.03</td>
<td>3.2e -14***</td>
</tr>
</tbody>
</table>

$t$, Time; DMT, Reading score. A positive estimate indicates that this predictor increases the amplitude of the ERP response (relative to the intercept), and a negative estimate indicates the opposite effect. The edf’s (estimated degrees of freedom) provide an estimation of the complexity of the nonlinear patterns being modeled (the higher, the more complex). Asterisks indicate significance: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Table 2.2: Random effects of the final model

<table>
<thead>
<tr>
<th></th>
<th>edf</th>
<th>Ref.df</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor smooth (t, Item × Hemisphere)</td>
<td>363.64</td>
<td>374</td>
<td>67.65</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>factor smooth (t, Subject × Hemisphere)</td>
<td>1328.94</td>
<td>2876</td>
<td>0.94</td>
<td>&lt;2e-16***</td>
</tr>
</tbody>
</table>

$t$, Time. The edf’s (estimated degrees of freedom) provide an estimation of the complexity of the nonlinear patterns being modeled (the higher, the more complex). Asterisks indicate significance: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

Interpreting fixed-effect predictors

Fig. 2.5: Visualizing the nonlinear interaction between Time and Reading score (DMT), separately for the Left (top) and the Right (bottom) hemisphere, and for Condition Symbol (Panel A), Condition Word (Panel B), and Word - Symbol difference (Panel C). The contour plots each shows a regression surface of the amplitude of the ERP signal as a joint function of Time (x-axis) and Reading score (y-axis). The contour lines represent isolines connecting combinations of Time and Reading score modeled to have the same amplitude. The colors correspond to the isolines, with blue, green, yellow and light grey indexing increasing values. Within the time window for the priming effect (between 175 and 275 ms), a positive relation between reading score and the amplitude of the word - symbol difference (i.e., the size of the tuning effect) is present in the left hemisphere (top row, Panel C), but not in the right hemisphere (bottom row, Panel C).
The nonlinear patterns modelled via the tensor product splines are best explained along with their visualizations. Fig. 2.5 displays the regression surface for symbols (Panel A) and for words (Panel B), separately for each hemisphere. In these contour plots, the amplitude of brain potentials is plotted as a joint function of time (ms; x-axis) and reading score (DMT; y-axis). The modelled amplitude of the ERP signal is color-coded, with blue, green, yellow and light grey indexing increasing values (see also the interpretation of Fig. 2.4 in the description of the Methods). Table 2.1 shows that all four regression surfaces plotted in Panel A and Panel B are highly significant (all $p$s < 0.001). The edf values (i.e., estimated degrees of freedom), which provide an estimation of the degree of nonlinearity, indicate that all the regression surfaces are highly nonlinear (all edf’s > 7).

Since the print-tuning effect is indexed by the difference between ERP response to words and that to symbols, we subtracted the regression surface for symbols from that for words (via the function `plot_diff2` from the R package `itsadug`; van Rij, Wieling, Baayen, & van Rijn, 2015), separately for each hemisphere. The resulting surfaces are presented in panel C. For the left hemisphere, between 175 and 275 ms, i.e., the time window for the N170 response, the word – symbol difference grows larger (more negative) as reading score increases, indicating a positive relation between the size of N170 print-tuning effect and reading performance. Such a pattern is missing in the same time window for the right hemisphere, where the unfolding of word – symbol difference follows a more complex, nonlinear pattern.

For clarification, Fig. 2.6 and Fig. 2.7 each offers a different type of visualization of the patterns shown in panel C of Fig. 2.5 (via the function `plot_diff` from the R package `itsadug`). In Fig. 2.6, time (x-axis) is plotted against the amplitude of the word – symbol difference (y-axis). For both hemispheres, Panel A visualizes the general pattern of the word – symbol difference over time, with reading score set to the average value (DMT = 52). Panels B, C, and D present time trends of the word – symbol difference for children with a reading score of 40, 60 and 80, respectively. The N170 print-tuning effect is clearly visible in the left hemisphere, as demonstrated by a prominent negative peak between 175 and 275 ms at all levels of reading performance. Consistent with the pattern observed in Fig. 2.5, this negativity indexing the N170 print-tuning effect grows larger as reading score increases from 40 to 60 and to 80. For the right hemisphere, however, visualizations of the difference wave over time do not support the existence of a robust N170: the word – symbol difference was not significant in children with average or lower reading scores, as the confidence bands largely overlap with the zero line in the first three plots. On the other hand, children with high reading scores (DMT = 80) did show a larger negativity to words than to symbols, nevertheless across the entire timespan under investigation (hence not specific to the N170), as reflected by an ascending difference wave (i.e., becoming more negative) in Panel D.

In Fig. 2.7, reading score (x-axis) is plotted against the amplitude of the word-symbol difference (y-axis), with Panels A to E showing the effect of reading score on the word – symbol difference at 175 ms, 200 ms (the median time point), 225 ms, 250 ms and 275 ms, respectively. For the left hemisphere, a positive, almost linear relation between reading score and the size of the print-tuning effect can be observed across the entire time window for the N170. For the right hemisphere, on the other hand, the effect of reading score is highly nonlinear: the word – symbol difference is reversed in polarity (i.e., positive, although not significant) for children with a moderate reading score (i.e., between 40 and 60), while a negative and significant word – symbol difference is found in children with a relatively high reading score (i.e.,
between 70 and 90). As shown in Panel D of Fig. 2.6, however, children with high reading scores showed a larger negativity to words than to symbols in the right hemisphere across the entire time span between 0 and 400 ms after stimulus onset. Therefore such enhanced negativity for words is not likely to index perceptual expertise specific to initial visual decoding of print stimuli.

Fig. 2.6: Visualizing the effect of Time on Word - Symbol difference for different reading scores (DMT), separately for the Left (top) and the Right (bottom) hemisphere. The amplitude of the word - symbol difference (y-axis) is plotted as a function of Time (x-axis). Panel A visualizes the unfolding of the word - symbol difference over time for children with an average reading score (DMT=52); Panels B, C, and D visualize the time trends for children with a reading score of 40, 60, and 80, respectively. For the left hemisphere (top row), within the time window for the print-tuning effect (between 175 ms and 275 ms), the amplitude of the word - symbol difference (i.e., the size of the tuning effect) grows larger as reading score increases. Such a positive relation is missing in the right hemisphere (bottom row).

**Interpreting random-effect predictors**

With respect to the random-effect factors, the effect of time varied significantly per subject and per item, indicating that there was large variability in the non-linear subject- and item-related ERP patterns (separately per hemisphere). We did not find support for between-condition variability in the individual time trends. This can be interpreted as words consistently eliciting a larger negativity than symbols across different subjects. In other words, the N170 tuning effect, as indexed by N170 amplitude-difference between words and symbols, can be robustly detected at the individual level.
2.4. Discussion

The objective of the current study was to investigate a possible neurophysiological marker of emerging dyslexia, i.e., visual specialization for print. Previous research has revealed deviant development of print-specific activation in beginning readers with dyslexia, as reflected by an absence of additional occipitotemporal activity for print (words or word-like letter strings) over visual baseline (symbols) in the dyslexic brain. At the neurophysiological level, this additional occipitotemporal activity is captured by the N170 component of the event-related potential. In the current study we investigated the relation between reading ability and the size, as well as the lateralization pattern, of the N170 print-tuning effect. For the first time, we used generalized additive modeling (GAM) to address this issue. This novel statistical tool characterizes the nonlinear shape of the ERP signal over the entire time window wherein the N170 tuning effect might emerge, thus yielding a more objective estimate of the effect than peak or mean amplitude measured from a pre-selected time window. Furthermore, by accommodating nonlinear interactions, GAM avoids arbitrary dichotomy of continuous variables (e.g., reading score), thereby decreasing the probability of finding spurious effects. Last but not least, GAM increases the

Fig. 2.7: Visualizing the effect of Reading score (DMT) on Word - Symbol difference at different time points within the time window for the print-tuning effect, separately for the Left (top) and Right (bottom) hemisphere. The amplitude of the word - symbol difference (y-axis) is plotted as a function of Reading score (x-axis). Panel A to E visualize the relation between reading score and the amplitude of the word - symbol difference at 175 ms, 200 ms, 225 ms, 250 ms, and 275 ms post stimulus onset, respectively. For the left hemisphere (top row), throughout the entire time window for the print-tuning effect, a positive, almost linear relation between reading score and the amplitude of the word - symbol difference (i.e., the size of the tuning effect) can be observed. Such a positive relation is missing in the right hemisphere (bottom row).
Chapter 2  Fast Visual Specialization for Print and Reading Fluency

generalizability of our findings by controlling for individual variations introduced by random-effect factors.

The results of our GAM analyses revealed a robust coarse tuning effect in the left hemisphere, as reflected by a prominent negativity peaking between 175 and 275 ms on the word – symbol difference wave. Crucially, the observed tuning effect was present across all levels of reading score, growing larger in size as reading score increases. In accordance with this pattern, visualizations of the effect of reading score at various time points demonstrated a positive, almost linear relation between reading score and the amplitude of the word – symbol difference across the entire time window for the N170 response. This set of findings, however, was limited to the left hemisphere. For the right hemisphere, the effect of reading score on the amplitude of the word – symbol difference was highly nonlinear. Only children with relatively high reading scores showed a reliably larger negativity to words than to symbols. However, such enhanced responses were present across the entire timespan under analysis (0 to 400 ms post stimulus onset), and therefore were not likely to reflect specifically initial visual decoding of print stimuli. Children with average or lower reading scores did not show a reliable word – symbol difference in the right hemisphere.

Our finding regarding the correlation between reading score and the size of the coarse tuning effect adds to existing evidence obtained using traditional ANOVA approaches. The coarse tuning effect, as indexed by an enhanced N170 response for words over visual baseline, has been found to develop rapidly during the initial stage of reading acquisition in typically developing children, but not in children with dyslexia (Araújo et al., 2012; Eberhard-Moscicka et al., 2015; Maurer et al., 2007, 2011). Using a nonlinear modelling approach (GAM), the current study expands upon earlier findings by characterizing the relation between reading fluency and the N170 coarse-tuning effect on a more fine-grained, continuous scale. Our results demonstrate that, not only does the presence of a coarse tuning effect differentiate between poor and normal readers, but the size of this tuning effect also correlates systematically with reading score. It is also worth noting that, in our study sample, words consistently elicited a larger negativity than symbols, as reflected by the lack of between-condition variability in individual ERP patterns over time (the factor smooth (Time, Subject × Condition) was not warranted in the random structure of the final model). This is in line with earlier findings that the coarse tuning effect could be robustly detected at the individual level (Eberhard-Moscicka et al., 2015). Taken together, these findings lend further support to the validity of the N170 coarse-tuning effect as a neurophysiological marker of dyslexia.

The correlation between the size of the print-tuning effect and reading ability may suggest a reciprocal relationship: on the one hand, the sensitivity for print only develops with increasing exposure to print; on the other hand, such sensitivity exerts strong influence on the outcome of learning to read. Specifically, print tuning is believed to index perceptual expertise for written language, which lies at the basis of fast, automatized visual word recognition that characterizes skilled reading. From a developmental perspective, visual decoding of print could start as part of the general object-recognition process, engaging an extensive network of visual processing mechanisms. With repetitive exposure to patterned orthographic information, mechanisms focusing on visual cues critical to letter identities are increasingly employed, and hence, tuned to print processing, while less appropriate mechanisms are gradually suppressed (Schlaggar & McCandliss, 2007). Supporting this perspective, functional imaging studies on the developing reading network have
demonstrated reduced activation in bilateral extrastriate regions, accompanied by increased involvement of the left ventral occipitotemporal cortex (Brown et al., 2005; Schlaggar et al., 2002; Shaywitz et al., 2002). Failure in this specialization process, as reflected neurophysiologically as reduced N170 tuning for print, leads to difficulties with visual word recognition, and ultimately, to problems with reading. There is evidence, however, that impaired visual specialization for print is best characterized as a developmental delay, rather than a persistent deficit. Following a group of dyslexic children and their age-matched controls, Maurer and colleagues (Maurer et al., 2011) found that the prominent coarse-tuning deficit characterizing dyslexic 2nd graders (Maurer et al., 2007) had largely disappeared by the 5th grade, presumably because visual word recognition has become automatized through years of reading practice, and most dyslexic children have caught up in this respect. Beyond the early years of reading acquisition, the coarse-tuning deficit seems to be restricted to severe cases of dyslexia (Helenius et al., 1999), and is most detectable when the reading system is seriously challenged, e.g., in explicit reading tasks (Maurer et al., 2011). Thereby, the N170 coarse-tuning deficit plays a particularly important role at the beginning of learning to read, and is best used as a diagnostic index for emerging dyslexia.

In our study, the N170 coarse-tuning effect was present only in the left hemisphere, as was the positive relation between reading score and the size of the tuning effect. This confirms previous results regarding the lateralization of the print N170, and accords well with the view that the left VWFS underpins visual specialization for print (Maurer et al., 2011, 2007; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999). Indeed, the left-lateralization pattern has been suggested to be an important feature of print tuning (Bentin et al., 1999), which differentiates visual expertise for language from that for general objects or faces (Maurer & McCandliss, 2007). According to the phonological mapping hypothesis, proposed by McCandliss and Noble (2003), the left-hemispheric modulation of the print N170 is induced by grapheme-to-phoneme conversion. Since the phonological and the orthographic systems are frequently co-activated for the task of reading, especially at the early stage of reading acquisition, difficulties with phonological processing in dyslexics could affect their visual-orthographic processing. Following this line of reasoning, the N170 print-tuning deficit, though occurring in the visual domain, might be related to the phonological-core deficit of dyslexia. In support of the phonological mapping hypothesis, a study by Brem and colleagues (Brem et al., 2013) found enhanced N170 responses following participation in a serious game designed to train grapheme-phoneme correspondences. An alternative account of the origin of the print N170, the perceptual account, relates the enhanced and left-lateralized N170 for print to visual familiarity with a particular script (e.g., Valdois et al., 2012). Under this framework, the print-tuning deficit is considered an independent visual processing deficit.

Whereas the coarse tuning effect can be robustly detected and systematically related to reading score, we did not find indication for a reliable fine tuning effect, as the children responded almost equally to words and to pseudowords across both hemispheres. This is not surprising given that the fine tuning effect doesn’t fully develop until reading has become automatic (McCandliss & Noble, 2003), and that the fine tuning effect, if developed at all, usually does not surface in implicit reading tasks, such as the repetition detection task used in the current study (Bentin et al., 1999; Maurer et al., 2005 a, b; Schendan et al., 1998; Wydell et al., 2003). As proposed by Eberhard-Moscicka and colleagues (2015), the lack of fine tuning effect in beginning readers might reflect their insufficient experience with written language.
A limitation of the current study is the subjectivity involved in the selection of the channels for analyses. Indeed, GAM can be used to model the topography of the ERP signals over the entire scalp, e.g., by creating a tensor product spline combining time and electrode position (coded with x and y coordinates). In future research, advances in the software would allow such complex modelling to take into account individual electrodes, thus offering a more complete representation of the data. Another limitation of our study is its small sample size. The correlation between reading score and the size of the N170 print-tuning effect needs to be tested with a larger sample in further studies.

To conclude, in the current study we used generalized additive modelling to investigate the nonlinear dynamics of neural activities underlying fast visual word recognition. Our results demonstrate a robust N170 coarse-tuning effect in the left hemisphere, as well as a positive relation between reading fluency and the size of the tuning effect. The N170 coarse-tuning effect, readily elicited by implicit reading tasks and robustly detected at the individual level, thus offers an early neurophysiological marker to distinguish poor from normal readers at the initial stage of reading acquisition.