

# Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG

O. Hauk\*, F. Pulvermüller, M. Ford, W.D. Marslen-Wilson, M.H. Davis

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK

## ARTICLE INFO

### Article history:

Received 3 October 2007

Accepted 18 April 2008

Available online 4 May 2008

### Keywords:

Written word recognition

Multiple linear regression

ERRC

Neighbourhood size

Word length

Orthographic typicality

## ABSTRACT

We applied multiple linear regression analysis to event-related electrophysiological responses to words and pseudowords in a visual lexical decision task, yielding event-related regression coefficients (ERRCs) instead of the traditional event-related potential (ERP) measure. Our main goal was to disentangle the earliest ERP effects of the length of letter strings (“word length”) and orthographic neighbourhood size (Coltheart’s “*N*”). With respect to *N*, existing evidence is still ambiguous with respect to whether effects of *N* reflect early access to lexico-semantic information, or whether they occur at later decision or verification stages. In the present study, we found distinct neurophysiological manifestations of both *N* and word length around 100 ms after word onset. Importantly, the effect of *N* distinguished between words and pseudowords, while the effect of word length did not. Minimum norm source estimation revealed the most dominant sources for word length in bilateral posterior brain areas for both words and pseudowords. For *N*, these sources were more left-lateralised and consistent with perisylvian brain areas, with activation peaks in temporal areas being more anterior for words compared to pseudowords. Our results support evidence for an effect of *N* at early and elementary stages of word recognition. We discuss the implications of these results for the time line of word recognition processes, and emphasise the value of ERRCs in combination with source analysis in psycholinguistic and cognitive brain research.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

The neuroscientific investigation of written word recognition faces the major problem that the processes of interest are affected by a large number of intercorrelated variables. The situation is complicated by the fact that highly correlated variables can affect different aspects of processing. For example, word length and orthographic neighbourhood size (i.e. the number *N* of words that can be obtained from a base word by exchanging just one letter, such as “can” into “car”), are negatively correlated in normal language. This reflects the fact that short words (such as “cat”) are commonly similar to more other words (such as “mat”, “fat”, “can”) than long ones (such as “crocodile”). While word length is commonly linked to early visual or orthographic processes (e.g. Ellis, 2004; Hauk and Pulvermüller, 2004a; Mechelli et al., 2000), effects of *N* have been interpreted in terms of competition during lexical access (Andrews, 1997; Grainger and Jacobs, 1996; Holcomb et al., 2002), or even post-lexical processing (Fiebach

et al., 2007). Thus new methods of disentangling these correlated variables are valuable if neuroscientists are to discover correlates of these psycholinguistic variables in early or late processes during the recognition of visually presented words.

Most electrophysiological and neuroimaging studies so far have been using factorial contrasts (e.g. words versus pseudowords, long words versus short words, etc.) in order to determine the effect of a particular variable on the brain response. This approach has two main disadvantages: (1) it does not exploit information in the continuous distribution of values, e.g. of word lengths for individual items. (2) In order to match for highly correlated confounding variables that are not of interest (e.g. for word length when effects of *N* are studied), “unusual” items on the extremes of the parameter distributions might have to be chosen (Baayen et al., 1997; Ford et al., 2003).

An alternative to factorial designs is multiple linear regression analysis. This method allows testing to what degree a variable (e.g. word length) predicts data across all trials. In the case of only one variable, this corresponds to computing the covariance between the data with the predictor variable. In the case of multiple intercorrelated variables, orthogonalisation procedures need to be applied in order to obtain independent estimates for each variable,

\* Corresponding author. Tel.: +44 1223 273702.

E-mail address: [olaf.hauk@mrc-cbu.cam.ac.uk](mailto:olaf.hauk@mrc-cbu.cam.ac.uk) (O. Hauk).

as will be described in more detail below. All other things being equal, regression designs will have greater power than dichotomizing continuous variables in a factorial experiment (Cohen, 1983).

Multiple linear regression analysis has long been applied to behavioural data on visual word recognition (e.g. Balota et al., 2004; Whaley, 1978). In neuroimaging, regression analysis is part of the commonly used “general linear model” (Friston et al., 1995), but it has not been widely applied in language research (see Davis et al., 2004; Graves et al., 2007, for examples). It has only very recently been introduced to the field of human electrophysiology (Dambacher et al., 2006; Hauk et al., 2006a). In the following, we will briefly summarise how electrophysiological studies on psycholinguistic variables have so far contributed to our knowledge about the early time course of visual word recognition. We will then motivate our interest in the particular variable *N*, which has been investigated by only few neuroscientific studies. We will also describe the multiple linear regression approach applied in the present study in more detail, emphasising aspects that are special to the analysis of electroencephalographic (EEG) or magnetoencephalographic (MEG) data.

So far, the most commonly assessed psycholinguistic properties are word frequency (estimate of the frequency of a word’s occurrence in the language) and lexicality (difference between words and non- or pseudowords); two variables which also produce reliable effects in behavioural data (Gernsbacher, 1984; Whaley, 1978). In electrophysiological research, Sereno et al. (1998) reported effects of word frequency on the “N1” component at 132 ms in a lexical decision experiment. A recent ERP study using lexical decision found word frequency effects interacting with emotional quality of words even earlier, around 100 ms (Scott et al., 2008). Effects of word frequency were detected slightly later by the studies of Hauk and Pulvermüller (2004a) around 160 ms in a lexical decision task, and Dambacher et al. (2006) around 170 ms for sentence reading. Assadollahi and Pulvermüller (2003) found an effect of word frequency in their MEG on word reading around 150 ms, that surfaced as an interaction with the variable word length. An interaction between word length and frequency was also found for words presented in sentence context around 120 and 180 ms (Penolazzi et al., 2007). An even earlier effect of word frequency occurred in the EEG study of Hauk et al. (2006b) around 110 ms, employing a lexical decision task. In all of these studies, higher frequency words generally evoked lower amplitude neural responses.

Lexicality effects have been consistently reported around 200–250 ms, for example as a “word recognition potential” (Hinojosa et al., 2004; Martin-Loeches et al., 1999; Rudell, 1991). Similarly, Dehaene (1995) reported differences between words and consonant strings at 192 ms. Cohen et al. (2000) described ERP differences between words and non-words shortly after 200 ms. Hauk et al. (2006b) reported a main effect of lexicality around 200 ms, but an earlier effect of lexicality that interacted with orthographic typicality around 160 ms. Hauk et al. (2006a) found a significant difference between words and pseudowords around 160 ms. The earliest lexicality effects so far have been reported by Sereno et al. (1998) at 100 ms.

This overview demonstrates that effects of lexicality and word frequency before 200 ms have occurred across several different studies. These results have been corroborated and extended by a few studies using linear regression or related methods. In the study of Hauk et al. (2006b), words and pseudowords were presented during an EEG experiment in a lexical decision task. The effects of several psycholinguistic variables on brain responses were evaluated using multiple linear regression on the EEG responses to words to derive event-related regression coefficients (ERRCs).

These ERRCs quantify the influence of specific parameters of written words on evoked electrophysiological responses (the regression equivalent of the “event-related potentials”, ERPs, derived from a traditional factorial design). Variables associated with the form of written words (word length and orthographic typicality as quantified by bi- and trigram frequencies) affected the ERRCs already around 90 ms after word onset. Effects of word frequency were detected shortly afterwards around 110 ms. The morpho-semantic variable Semantic Coherence (describing the consistency of meanings within a morphological family; Ford et al., 2003; Landauer and Dumais, 1997) was reflected in the ERRCs around 160 ms, co-occurring with the first significant difference between words and pseudowords.

This pattern of results was interpreted in terms of a serial but cascaded sequence of processing steps within the first 200 ms after presentation of a written word. This interpretation was further supported by source estimation results. Length produced right-lateralised activation in posterior brain areas consistent with basic visual analysis of the word form. Activation for orthographic typicality was left-lateralised, and the largest centre of activation occurred at a left inferior temporal location. The earliest influence of word frequency was also left-lateralised, and located in a posterior temporal area. The source estimates reflecting Semantic Coherence, as well as for the word/pseudoword differences, were more distributed across both hemispheres. Interestingly, around 200 ms as well as at later latencies between 300 and 500 ms, several psycholinguistic variables produced effects simultaneously and with similar topographies. This lack of specificity may indicate either integration of information across psycholinguistic processing levels, or post-access verification and decision processes.

Similar evidence for early lexical access for written words have also been obtained for sentence stimuli, Dambacher et al. (2006) presented sentences word-by-word to their subjects, and varied the lexical frequency of open-class words at different positions within the sentence, as well as their predictability by the context. Regression analysis revealed a context-independent effect (i.e. not depending on word position within a sentence) of word frequency around 170 ms after word presentation. The N400 component was affected by both predictability and word frequency. They suggested that lexical information is accessed before 200 ms, and that later ERP components reflect contextual integration processes. Dien et al. (2003) used items-analysis in their analysis of EEG responses to congruous or incongruous sentence endings, respectively. They averaged epochs for individual stimulus items across subjects, and applied parametric analysis to the resulting data set. The variables of interest were ‘meaningfulness’ and ‘expectedness’ of the target word (i.e. how much sense the sentence makes including the target words, versus how strongly the participants expect this word given the preceding context). Earliest effects of both variables were detected around 200 ms after word onset, suggesting that lexical information about both the target word and the context in which it occurs is already available at this latency. This is consistent with findings that effects of word class (e.g. verbs versus nouns) or semantic and emotional attributes (for example effector type of actions) have been reported around 200–250 ms (Hauk and Pulvermüller, 2004b; Kissler et al., 2008; Pulvermüller et al., 1995; Pulvermüller et al., 1999; Skrandies, 1998).

Another variable that potentially taps into lexical processing, but has only been investigated in a few electrophysiological and neuroimaging studies, is orthographic neighbourhood size, or *N* (Coltheart et al., 1977). This variable is of particular interest because it reflects the orthographic relatedness of a letter string with words in memory, and might therefore affect competition and inhibition processes in word retrieval. *N* can be computed for both

words and pseudowords (in contrast to word frequency, for example), and indeed effects of  $N$  in behavioural tasks have been reported to differ between words and pseudowords. In the original lexical decision study of Coltheart et al. (1977), effects of  $N$  were reported only for nonwords, and were inhibitory (i.e. slower responses to more word-like nonwords which have a higher  $N$ ). More recent studies have confirmed that rejections for pseudowords are slower for higher  $N$ 's, but have also found facilitatory  $N$  effects for words (Andrews, 1989; Forster and Shen, 1996; Grainger and Jacobs, 1996; Sears et al., 1995). This pattern has also been found in the behavioural data of Holcomb et al. (2002), while their ERP data showed the same increase of N400 amplitudes with  $N$  for both words and pseudowords.

Opinions differ with respect to the origin of these orthographic neighbourhood effects. It has been suggested that  $N$  facilitates the lexical retrieval process at an early stage by means of feedback from the lexical to the letter level (Andrews, 1997), and that pseudowords with high  $N$  are more difficult to reject because they at least partly activate the corresponding word neighbours (Sears et al., 1999). In the multiple read-out model of Grainger and Jacobs (1996), it is assumed that lexical decisions are based on summed activation across all activated word representations. If a letter string activates many neighbours, the corresponding summed activation will be higher, which facilitates responses to words, but slows down rejections for pseudowords. In tasks where decisions are based on activation of individual representations, such as in word identification, the activation of neighbours inhibits the selection of the target item, and responses are therefore slower for higher  $N$ s (e.g. Perea et al., 2004). This interpretation has been challenged by other researchers, who found  $N$  effects to depend on other task and stimulus properties. For example,  $N$  effects were not found for “No” responses in a semantic classification task (Forster and Shen, 1996), and have been reported to depend on the matching between word and non-word stimuli (Siakaluk et al., 2002). It is therefore still a matter of debate whether effects of  $N$  reflect fundamental lexical selection processes, or rather task-specific response strategies (see Andrews, 1997; Balota et al., 2004; Norris, 2006, for overviews).

Determining the time course and neuronal correlates of  $N$  effects in evoked brain responses may contribute to the understanding of processes modulated by orthographic neighbourhood variables. Early effects of this variable, i.e. in the latency range of the earliest previously reported effects of orthographic and lexical variables, would indicate that it indeed affects elementary word recognition processes. This argument would be strengthened if brain activation occurred in “classical” left-lateralised perisylvian language-related brain areas. Currently available neuroimaging data on  $N$  are as yet inconsistent. Binder et al. (2003), using a visual lexical decision task and orthographically matched words and pseudowords, did not find any brain areas for which activation significantly increased with neighbourhood size. Instead, they found that higher  $N$  produced lower activation to words in left prefrontal, angular and ventrolateral temporal cortex. In contrast, Fiebach et al. (2007) found differential effects of  $N$  for words and pseudowords in a lexical decision task in medial and mid-dorsolateral prefrontal cortex. Because these areas are commonly related to executive control functions rather than lexico-semantic processing, the authors argue that effects of  $N$  might arise only at a late post-lexical level. Consistent with this suggestion, the EEG study of Holcomb et al. (2002) revealed  $N$  effects around 400 ms after stimulus onset in a lexical decision task, which could reflect a later processing stage. In a semantic categorisation task, however, effects occurred earlier between 150–300 ms. We conclude at this point that effects of  $N$  are of great interest for psycholinguistic theories of lexical access, and that more data are needed to

establish its effect on brain activation, and in particular its time course.

The investigation of  $N$  is complicated by the fact that  $N$  is negatively correlated with word length (e.g. Weekes, 1997). It is also positively correlated with typicality, i.e. more typically spelled words have more orthographic neighbours. The fMRI studies cited above controlled for word length and, where applicable, word frequency, but not orthographic typicality (Binder et al., 2003; Fiebach et al., 2007). Only the study of Holcomb et al. (2002) controlled all three of these variables. In our previous study (Hauk et al., 2006a), word length was negatively correlated with orthographic neighbourhood size and both variables were combined into a single predictor variable. Furthermore, results from the regression analysis were only presented for word stimuli, but not for pseudowords. In the current paper, we will therefore present a new analysis of the data reported by Hauk and colleagues, focusing on three novel questions: (1) Can we further characterise early brain responses by entering separate variables for  $N$  and word length into the regression analysis? (2) Can we reveal early differences between words and pseudowords for these variables, similar to those reported in the behavioural literature? (3) Do the source distributions differ between words and pseudowords, and how do they compare to existing neuroimaging results? In order to address these questions, we performed a new analysis on pseudoword data, and a re-analysis of our word data, including three variables: (i) word length measured as number of letters; (ii) orthographic neighbourhood size ( $N$ ); (iii) orthographic typicality measured by bi- and trigram frequencies. Because converging results for the variable Typicality have already been presented in two independent studies (Hauk et al., 2006a,b), we will focus our analysis on the variables  $N$  and word length.

## 2. Methods: general

Before describing the specific methodological setup of the present study, we will present some general information about multiple regression and its combination with source estimation.

### 2.1. Multiple linear regression

Recently, multiple linear regression has been used in order to determine the impact of several psycholinguistic variables – tapping into different stages of word processing – on EEG amplitudes in a single experiment (Hauk et al., 2006a). The resulting measure was called ERRCs, in contrast to event-related potentials (ERPs). In the case of multiple intercorrelated variables, the solution of multiple regression provides a best least-squares fit of all variables simultaneously to the data (Bertero et al., 1985; Howell, 2001), i.e. it is the least-squares solution of the linear equation:

$$\mathbf{d} = \mathbf{X}\mathbf{b}$$

where  $\mathbf{d}$  is a vector representing the measured data (with dimension  $n$ , e.g. ERP amplitudes at one electrode and latency after stimulus onset across all  $n$  events),  $\mathbf{b}$  a vector with  $m$  elements representing regression coefficients for the  $m$  predictor variables (i.e. length, neighbourhood size, and typicality in the present study), and  $\mathbf{X}$  is a matrix with  $n$  rows and  $m$  columns containing the predictor variables as columns (i.e. each column contains the values for length,  $N$ , and typicality for the corresponding event). The unique solution of this equation in the least-squares sense is obtained by the generalised inverse matrix (e.g. Golub and van Loan, 1996):

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{X}^- \mathbf{y}$$

where  $\mathbf{X}^T$  represents the transpose of the matrix  $\mathbf{X}$ , and  $\mathbf{X}^-$  the generalised inverse of the matrix  $\mathbf{X}$ , both with dimension  $m \times n$ . Each row of  $\mathbf{X}^-$  is associated with one predictor variable (e.g. length), such that it is correlated with this variable, but orthogonal to all others (more precisely:  $\mathbf{X}^- \mathbf{X} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix) (Golub and van Loan, 1996). This means that multiple linear regression yields a weighted average of the data for each predictor variable, with the important property that the new variables in  $\mathbf{X}^-$  are decorrelated with respect to the predictor variables in  $\mathbf{X}$ , and therefore yield independent estimates for each of them. It is interesting to note, as was pointed out previously (Hauk et al., 2006a), that factorial designs are a special case of linear regression: instead of weighting the data with values from a continuous predictor variable (or the generalised inverse), they are weighted by  $1/n$  in one condition, and  $-1/n$  in another ( $n$  being the number of events that enter the corresponding averages), resulting in the subtraction of the

averages of these two conditions. Linear regression analysis therefore does not require more processing on the single-trial level than classical averaging.

A common problem with general linear models is the scaling of the predictor variables. It is usually meaningless to compare results for variables that are measured in different units, such as number of letters versus number of occurrences per million in a database. Therefore, comparisons should be made only for each variable separately (e.g. for word length at a particular latency versus baseline), or appropriate scaling procedures have to be applied. In the study of Hauk et al. (2006a), the former approach was chosen, i.e. latencies at which signals exceeded the baseline level were determined for individual predictor variables separately. Furthermore, topographies were compared between variables at selected latencies after data were normalised for overall amplitude, i.e. after removing differences due to scaling of the variables. In order to compare regression coefficients obtained for different stimulus groups (such as for words and pseudowords in the present study), it must be assured that data are scaled equally for both groups, e.g. that all predictor variables are unitless. Generally, including a scaling matrix into the initial linear model above yields

$$\mathbf{d} = \mathbf{X}(\mathbf{W}\mathbf{b}) = (\mathbf{X}\mathbf{W})\mathbf{b}$$

This, after some matrix calculus, results in the new solution:

$$\mathbf{b} = \mathbf{W}^{-1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{W}^{-1}\mathbf{X}^{-\mathbf{y}}$$

if  $\mathbf{W}$  is a symmetrical and invertible matrix. This means if each predictor variable was scaled by an individual value (i.e.  $\mathbf{W}$  is a diagonal weighting matrix), the result for each regression coefficient is the outcome of the unweighted equation divided by the weighted coefficient for the corresponding variable. If the result is supposed to be rescaled into its original units (for example microvolts), it has to be multiplied by the corresponding weighting coefficients.

Multiple linear regression, in the framework of general linear models, have already been applied in a number of fMRI studies (see Friston et al., 1995 for the theoretical foundation, and Graves et al., 2007 for an example in picture naming). A few studies have recently applied this or similar approaches to the analysis of EEG and MEG data (Dambacher et al., 2006; Dien et al., 2003; Hauk et al., 2006a).

## 2.2. Source estimation

Source estimation has been widely used in the analysis of evoked responses of both EEG and MEG data. We will here summarise the dominant approaches and highlight the important issue for linear regression analysis. The signal measured at or above the scalp (even if measured with perfect accuracy) alone does not provide enough information in order to uniquely determine the underlying generators (Fokas et al., 2004; Hämäläinen and Ilmoniemi, 1984; Hauk, 2004; Ilmoniemi, 1993; Sarvas, 1987). The only way of finding a unique solution is to impose further constraints independently of the measured data. Because the physiology of the generators and the mathematical formulation of the inverse problem are the same for both EEG and MEG, the same source estimation methods can be applied to both methods as well as their combination.

One approach widely employed is equivalent dipole modelling (e.g. Scherg, 1992). However, dipole modelling can only be expected to produce accurate results if the underlying modelling assumptions are correct. Crucially, the number of distinct activated brain areas has to be known in advance. In experiments that go beyond early sensory processes, however, this is usually an unknown which one would like to determine from the data. This problem is addressed by “distributed source models” (e.g. Fuchs et al., 1999; Michel et al., 2004). The continuous brain volume or the cortical surface is approximated by a large number of discrete densely spaced dipole sources, e.g. on a grid or a net. The strengths and orientations of these source is calculated under assumptions with respect to global properties of the distribution, e.g. minimising overall source strength (“minimum norm” approach), maximising a measure of likelihood (“Bayesian” approach), or creating a set of optimal “spatial filters” (“beamforming” approach). Although the terminology used to motivate and describe these different approaches differs significantly, their mathematical formulation can be very similar or even equivalent (Hauk, 2004). It is important to realise that no matter what approach is chosen, one cannot extract information from data that is missing. Non-uniqueness is inherent in the bioelectromagnetic inverse problem, and cannot be remedied by sophisticated mathematical algorithms. One either has to add reliable information at the beginning, by incorporating constraints into the algorithms, as with dipole models; or one has to choose an approach that relies on minimal modelling assumptions, and take into account the general resolution limits at the interpretation stage, as for distributed source models.

Source estimation is usually applied to event-related potentials or fields, i.e. after averaging trials within stimulus categories. This can be performed on grand-mean data, in order to enhance the signal-to-noise ratio before source analysis. However, the result of the combination of linear methods does not differ, by definition, on the sequence in which they are applied to the data. Because averaging across trials or subjects is a linear operation, linear source estimation procedures can be applied to the averaged data without loss of information, which reduces computation time. If one step of the analysis is non-linear, one has to choose the sequence of analysis steps more carefully. For example, when power spectra or measures of coherence

are computed on single-trial basis, source estimation afterwards would not produce meaningful results. In this case, it has to be applied on the single-trial level before the non-linear transformation. This already rules out methods that rely on restrictive modelling assumptions or that are highly sensitive to noise, such as equivalent dipole models. In this case, distributed source models have to be used (Gruber et al., 2006; Hauk et al., 2002).

Multiple linear regression is, by definition, a linear transformation of the data. Therefore, linear source estimation can be applied to the resulting regression coefficients, rather than on each single trial. However, linear source estimation is only linear as long as the orientation of sources is taken into account. As soon as the three components of a dipole source are combined into one source strength, i.e. into intensity values, the transformation becomes non-linear (this is similar to “rectifying” a bipolar signal). In many classical ERP or ERF studies, source orientation is not given much attention, and the main parameters are source location and strength. In factorial designs, it is usually the main concern whether one condition A produces larger amplitudes than another B in a particular brain area. This can be addressed by computing source intensity distributions for the two conditions separately and then subtracting them (A – B). In this case, positive difference values indicate that A produces more activation than B, and negative values the opposite, without information about the orientation of sources. Although this approach is justifiable, it is important to note that the difference of these intensity maps is not necessarily the same as the one obtained by computing intensity maps for single trials before averaging. Similarly, intensity maps computed on grand-mean ERP or ERF data are not necessarily the same as those computed as averages across intensity maps for individual subjects (Hauk et al., 2003). Computing intensity maps after averaging is justified only if one can assume source orientations to be relatively stable across subjects or trials, respectively.

In the case of linear regression, computing intensity distributions after linear regression removes the information about the sign of the regression coefficients. This procedure can determine in which brain areas activation covaries most strongly with a predictor variable, but not whether the regression coefficients are positive or negative. If this information is required, one would have to know not only the orientation of the source for the regression result, but also the orientation of the source for the average across all items. This is the same ambiguity with respect to difference amplitudes and polarity in the ERP signal: if a stimulus set is split into two groups A (high values for variable V) and B (low values for V), then a positive value for A – B or a positive covariance with V can be interpreted as more positive-going ERPs for high values of V, or more negative-going ERPs for low values of V. An alternative approach is to compute intensity maps for single trials, and applying regression analysis across these trials afterwards. The result would indicate whether the intensity of brain activity in a particular brain area is positively or negatively correlated with a particular variable, irrespective of its orientation. Depending on the number of source locations included into the model, this approach can be computationally demanding.

## 3. Methods: experiment

### 3.1. Subjects

Data sets from 20 right-handed monolingual native speakers of British English entered the final analysis (11 female, 9 male; mean age 22 years, S.D. 3; at least 14 years of school and higher education). All had normal or corrected-to-normal vision and reported no history of neurological illness or drug abuse. Handedness was determined according to a simplified version of Oldfield’s handedness inventory (Oldfield, 1971), revealing a mean laterality quotient of 85 (S.D. 25). Five subjects were initially removed from the data set due to extensive systematic eye blinking or other artefacts. Informed consent was obtained from all subjects and they were paid for their participation. This study was approved by the Cambridge Psychology Research Ethics Committee.

### 3.2. Stimuli and procedure

Three hundred monomorphemic English nouns were selected that were either lexically unambiguous nouns or, if lexically ambiguous, were used much more frequently as nouns than as members of other lexical categories (mean noun/verb frequency ratio, 22/1, CELEX database; Baayen et al., 1993). Homophonic words were excluded by checking all words in the Wordsmyth on-line dictionary (Parks et al., 1998). All words were between three and six characters in length and most were monosyllabic. Three hundred pseudowords were created according to the orthographic and phonotactic rules of British English. Those were matched for length and bigram frequency to the real words. Pseudowords did not include letters or letter combinations that could be interpreted as prefixes or suffixes. The average frequency of its letter bigrams and trigrams was calculated for each letter string along with word length counted in number of letters. Furthermore, the number of lexical neighbours was used to estimate the orthographic neighbourhood density (Coltheart’s N) (Coltheart et al., 1977). Means and standard deviations for the variables length, bigram and trigram frequencies, and N are presented in Table 1.

N and length (number of letters) entered the multiple regression design directly, while bigram and trigram frequencies were combined by computing their first principal component. This was considered appropriate since these variables are

**Table 1**

Means and standard deviations for the variables that were used in the multiple linear regression analysis

	# letters	N	Bigrams	Trigrams
Mean				
Words	4.40	7.05	38,118	4651
Pseudowords	4.39	7.07	37,849	4634
S.D.				
Words	0.86	5.42	15,824	3890
Pseudowords	0.88	5.40	15,660	3887

# letters: number of letters; N: Coltheart's orthographic neighbourhood size; bigrams/trigrams: bigram and trigram frequencies per million.

highly correlated and differential effects of these two variables were not of interest for the present study. The resulting three variables were z-normalised before entering the regression analysis. Note that, because standard deviations are matched for each variable between word and pseudoword stimuli, scaling does not affect the corresponding results. High correlation between variables in a multiple linear regression design can lead to numerical instabilities and increased sensitivity to noise. One way to quantify the degree to which the linear system is “ill-conditioned” is to compute the condition number, i.e. the ratio between the largest and the smallest eigenvalue of the matrix containing the predictor variables (Bertero et al., 1988), which can range from 1 (all eigenvectors contribute equally to variance) to infinity (only one eigenvector contributes to variance). For our stimulus set, the condition numbers for word and pseudoword stimuli were 2.46 and 2.20, respectively.

Participants performed a lexical decision task. White letter strings were presented on a grey background on a computer screen. Each stimulus was presented for 100 ms. The stimulus onset asynchrony (SOA) varied randomly between 2.5 and 3 s. A fixation cross was shown in the centre of the screen when no letter strings were present. Subjects were instructed to press one button of a response box with the index finger of their left hand in response to a real word, and another button with the middle finger of the same hand in response to a pseudoword, while minimising any other body and eye movements throughout the experiment. They were given several practice trials until they felt comfortable with the task. Stimulus sequences were randomised across subjects. The stimulus delivery and response collection was controlled by the Experimental Run Time System software (ERTS, BeriSoft, Germany).

### 3.3. Data recording and pre-processing of EEG data

The EEG was measured in an electrically and acoustically shielded EEG chamber at the MRC Cognition and Brain Sciences Unit in Cambridge, UK. Data were recorded from 63 unipolar EEG and 2 bipolar EOG electrodes (Ag/AgCl), all of which were mounted on an electrode cap (EasyCap, Falk Minow Services, Herrsching-Breitbrunn, Germany) except the lower vertical EOG electrode which was placed below the right eye, using SynAmps amplifiers (NeuroScan Labs, Sterling, USA). Electrodes were arranged according to the extended 10/20 system. Data were sampled at 500 Hz with a band-pass filter 0.1–100 Hz, and off-line band-pass-filtered between 1–20 Hz. Cz was used as recording reference for the EEG channels. The EOG was recorded bipolarly through electrodes placed above and below the left eye (vertical) and at the outer canthi (horizontal). The continuously recorded data were divided into epochs of 800 ms length, starting 100 ms before stimulus onset. Trials with peak-to-peak potential differences larger than 100  $\mu$ V in at least one EEG or EOG channel were rejected, as were trials in which incorrect responses were given. For each channel the mean amplitude of a 100 ms baseline interval was subtracted at all time points, and data were converted to average reference.

### 3.4. ERRC analysis

ERRCs were computed for a multiple linear regression design consisting of N, length and typicality, applied to word and pseudoword data separately for each subject ( $n = 20$ ) and each electrode ( $n = 63$ ). Our analysis was guided by the results obtained previously for the word data (Hauk et al., 2006a). The earliest effects of psycholinguistic variables were obtained around 100 ms, starting with word length and typicality. Peaks in the root-mean-square (RMS) curves appeared in the average across all words at 114, 160, 202, 314 and 500 ms, respectively. For pseudowords, these latencies were 116, 160, 202, 320 and 552 ms. Because of these similarities, we chose the earliest time windows from our previous study also for this analysis, i.e. 80–100, 100–120, 140–180 and 202–222 ms. In order to compare our results to some previous N400 studies, we also selected a time range 400–600 ms for analysis.

Our statistical analysis followed our earlier approach very closely (Hauk et al., 2006a). *t*-Tests were computed for each time window at each electrode, and the resulting *p*-value distribution plotted as topographic maps (see Figs. 2 and 3). This can reveal whether the most dominant effects reflect the most salient features of the ERRC topographies (i.e. coincide with maxima or minima, for example), and

whether they are captured by the selection of peak electrodes (see below). We also subjected our data to a more conventional statistical analysis. Topographies for word and pseudoword stimuli were very similar (see Fig. 2), in particular electrodes that showed maxima (positive or negative) in the topographical ERP distributions were also the same for averages across all words and all pseudowords, respectively. For that reason, we used the same nine electrodes as in the previous study: F7, Fz, F8; T7, Cz, T8; P7, Pz, P8. These were grouped into the factors Gradient (anterior–posterior, three levels) and Laterality (left–right, three levels) for analysis of variance (ANOVA). This resulted in a design with the factors Lexicality (words versus pseudowords), Laterality and Gradient, which was applied to data for each psycholinguistic variable separately. We applied Greenhouse–Geisser correction to the degrees of freedom where appropriate.

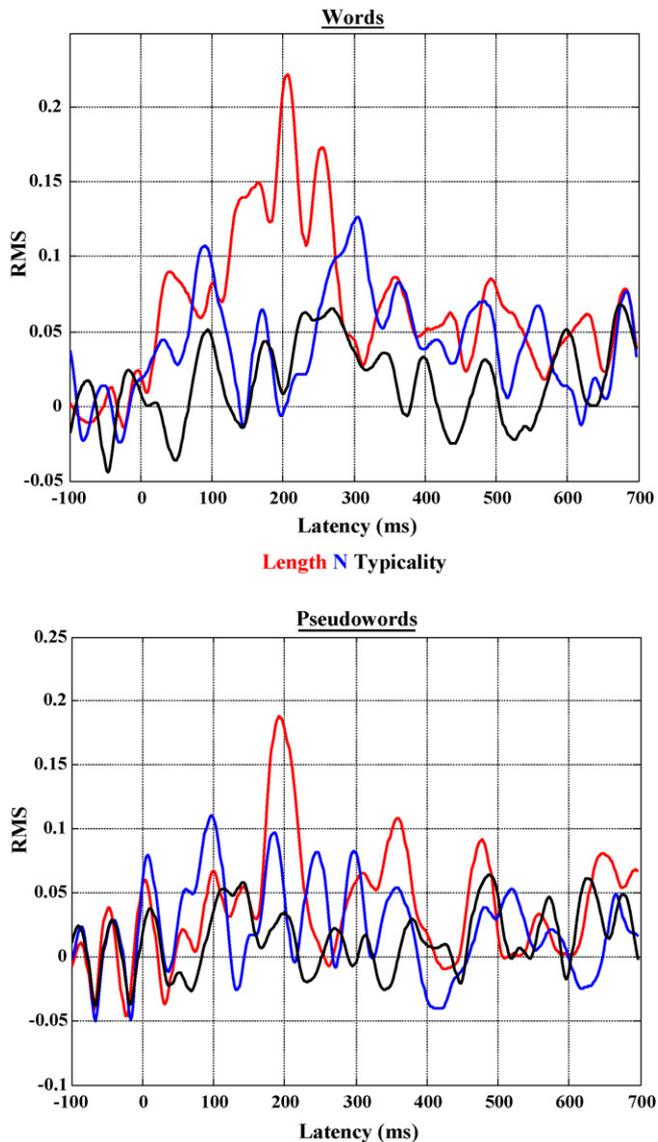
Minimum norm source estimation was applied following the approach of Hauk (2004). In the case of noiseless data, this method produces the unique solution among the infinitely many possible ones that explains the data completely, but has minimal overall source strength in the least-squares sense (Bertero et al., 1985; Hämäläinen and Ilmoniemi, 1984). The solution does not contain any “silent sources”, i.e. sources that do not produce any measurable signal at the recording electrodes (Hämäläinen and Ilmoniemi, 1984; Hauk, 2004). We applied minimum norm estimation on a two-dimensional source space, created as the envelope of the brain compartment of the standard brain of the Montreal Neurological Institute. The forward solution was computed using the boundary element method (BEM) in the Curry software (Neuroscan Corporation). This method yields a blurred two-dimensional projection of the true source distribution within the brain. The purpose of this analysis was to estimate possible generators for the significant effects revealed by our ERRC analysis. We therefore applied this method to our grand-mean data for different variables.

## 4. Results

We analysed the time ranges 80–100, 100–120, 140–180, 202–222 and 400–600 ms, for which ERRC topographies and *t*-test statistics are presented in Figs. 2 and 3. Data from nine peak electrodes were subjected to an ANOVA including the factors Lexicality (words/pseudowords) and the topographical factors Gradient (anterior/posterior) and Laterality (left/right) for N and Length separately. Fig. 1 illustrates the time course of ERRCs for the different psycholinguistic variables. For both words and pseudowords, the RMS curves for the different variables exceed the baseline level around 100 ms after stimulus onset. Note that the RMS collapses data across all electrodes, and therefore effects that are present at only few electrodes might be attenuated. Topographic maps, as presented in Figs. 2–4, are more informative about the location of effects. Fig. 2 shows grand-mean ERP topographies for averages across all words and all pseudowords, respectively. These maps illustrate that the main features of the spatial distribution of the ERPs are very similar for both stimulus categories, and that therefore the choice of the same latency ranges and peak electrodes for statistical analysis is justified.

Figs. 3 and 4 present the ERRC topographies for the variables Length and N in more detail. For Length, ERRC distributions in the earliest time ranges around 90 ms and 110 ms (Fig. 2) are characterised by bilateral posterior positive peaks for both words and pseudowords, reflecting a positive covariation of this variable with the brain response. Because the P100 ERP component exhibits positive values at posterior electrode sites (see Fig. 2), this means that greater length of the letter string predicts larger P100 amplitudes, in accordance with previous studies (Assadollahi and Pulvermüller, 2003; Hauk et al., 2006a; Hauk and Pulvermüller, 2004a). The topographical ANOVA did not reveal effects of Lexicality for the variable Length in these two latency ranges, i.e. effects of Length were statistically indistinguishable for words and pseudowords. However, a main effect of the factor Lexicality occurred for Length between 140 and 180 ms ( $F(1, 20) = 6.07$ ,  $p < 0.05$ ). Figs. 1 and 3 illustrate that this is due to larger ERRCs associated with Length for words compared to pseudowords. The largest effects of Length occurred around 212 ms, and again did not distinguish between words and pseudowords.

ERRC topographies for the variable Neighbourhood Size showed different topographies for words and pseudowords already in the



**Fig. 1.** Root-mean-square (RMS) curves of event-related regression coefficients (ERRCs) for words (top) and pseudowords (bottom). Curves are plotted separately for the psycholinguistic variables 'Length' (red), 'Neighbourhood size' (blue), and 'Orthographic typicality' (black).

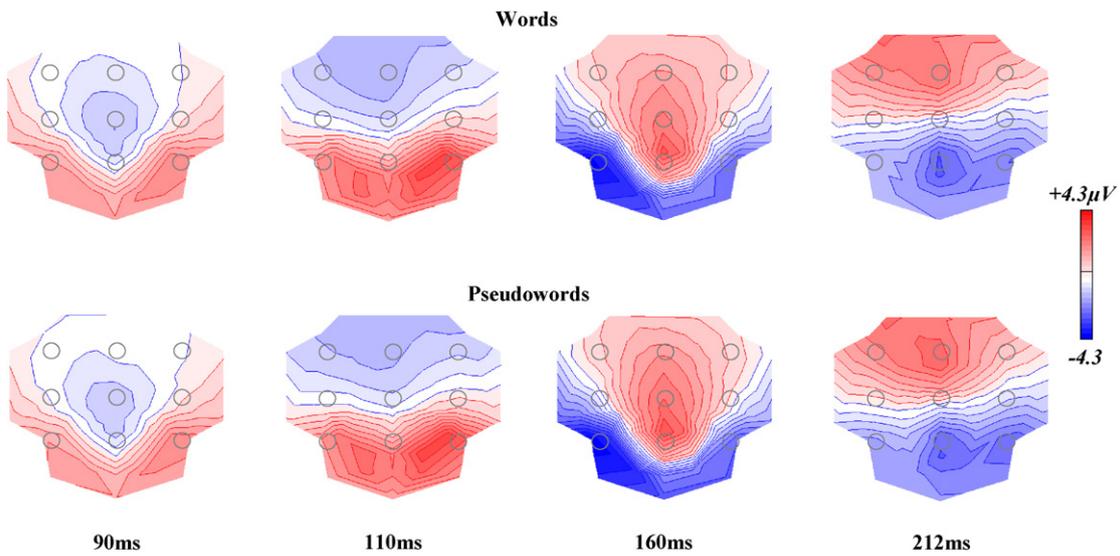
latency ranges 90 ms and 110 ms (Fig. 4). An interaction Lexicality-by-Gradient-by-Laterality was detected already between 80 and 100 ms ( $F(4, 80) = 3.06, p < 0.05, \epsilon = 0.59$ ), as well as an interaction Lexicality-by-Laterality ( $F(2, 40) = 9.2, p < 0.001, \epsilon = 1.0$ ). A topographical interaction with the factor Laterality was also found in the time range 100–120 ms ( $F(2, 40) = 3.31, p < 0.06, \epsilon = 0.86$ ), while the three-way interaction was not significant for this time range ( $F(4, 80) = 1.61, p > 0.1$ ). Fig. 4 shows that effects of  $N$  for words are characterised by negative ERRCs at left-frontal and positive ERRCs at centro-parietal electrode sites, while pseudowords exhibit positive ERRCs at left-posterior and negative ERRCs at fronto-central electrodes. Similar to the effects of Length reported above, the positive covariation of  $N$  with positive ERP amplitudes at posterior electrodes (see Fig. 2) demonstrates that more orthographic neighbours are reflected by higher absolute amplitudes. No further effects of the factor Lexicality were found in the remaining time ranges (all  $F < 2.1, p > 0.5$ ), although in the time range 400–600 ms the interaction Lexicality-by-Gradient approached significance ( $F(2, 40) = 3.24, p = 0.08, \epsilon = 0.58$ ).

Fig. 5 presents minimum norm source estimates for selected time ranges. Our main interest was in the earliest detectable effects of Length and  $N$ . ERP topographies for each of these variables were very similar for words and pseudowords in the latency ranges 80–100 and 100–120 ms. We therefore combined ERP data for these two ranges into one time window 80–120 ms before applying source estimation, in order to enhance the signal-to-noise ratio. Because no significant difference between words and pseudowords was found for Length in the ERRC data, we computed minimum norm source estimates for data collapsed across words and pseudowords. The source estimates for Length in Fig. 4 exhibit two prominent activation centres in posterior brain areas, with larger activation in the right hemisphere. This replicates the findings of Hauk et al. (2006b), and extends them to pseudowords. In the case of pseudowords, source estimation results for  $N$  resemble those obtained for Length: the main activation centres were found bilaterally in posterior brain areas. However, for  $N$  this activation was stronger in the left hemisphere. A smaller activation peak, which was not visible for Length, was present in a centro-parietal area. In the word data,  $N$  did not produce clear activation peaks in the right hemisphere. The activation centre in the left hemisphere was more anterior and inferior to the one found in the pseudoword data. A centro-parietal activation peak was also present for words.

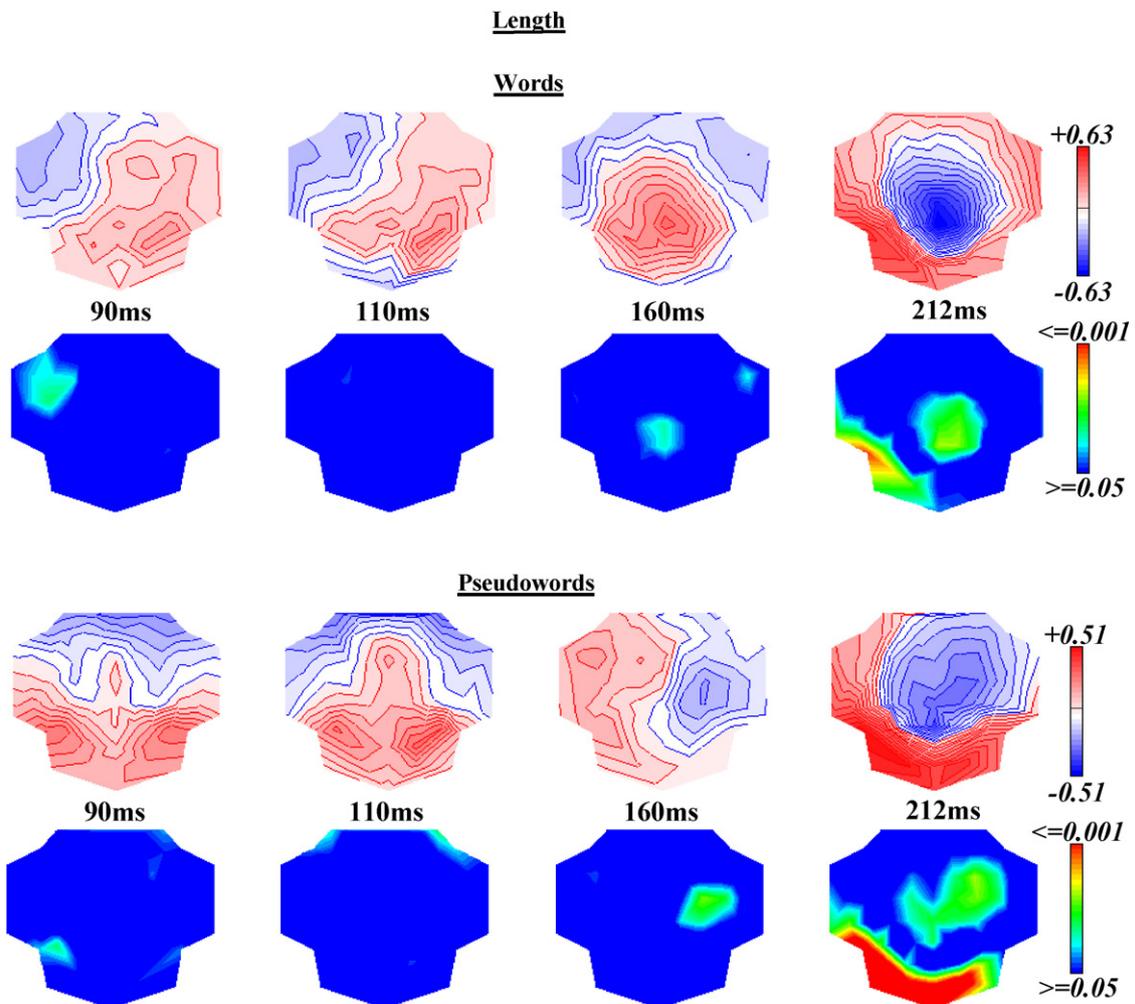
## 5. Discussion

Our analysis confirmed previous results that psycholinguistic variables affect electrophysiological brain responses already within the first 200 ms (see above). We found ERRC effects of word length around 100 ms after word onset, which correspond well with those reported in previous studies (Assadollahi and Pulvermüller, 2003; Hauk and Pulvermüller, 2004a), and in particular with the results of our previous analysis (Hauk et al., 2006a). Our multiple regression approach allowed us to analyse effects of neighbourhood size ( $N$ ) and word length for the same subjects and items. We found the earliest effects of  $N$  around 100 ms. Furthermore, our study is the first to report that effects of  $N$  depend on the lexicality of the letter strings: covariations between  $N$  and neural activity were present for both words and pseudowords, but differed in topography. This indicates that the brain response at 100 ms is already sensitive to whether a letter string has been previously experienced or not. Interestingly, our analysis revealed that the ERP amplitude in the P1 latency range increased with  $N$ . This is consistent with the view that letter strings partly activate memory traces for their orthographic neighbours during the word retrieval process, both for words and pseudowords. The fact that pseudowords do not have memory traces themselves, but only activate memory traces of their neighbours, should cause differences in the retrieval process that might underlie the topographical differences observed in our study.

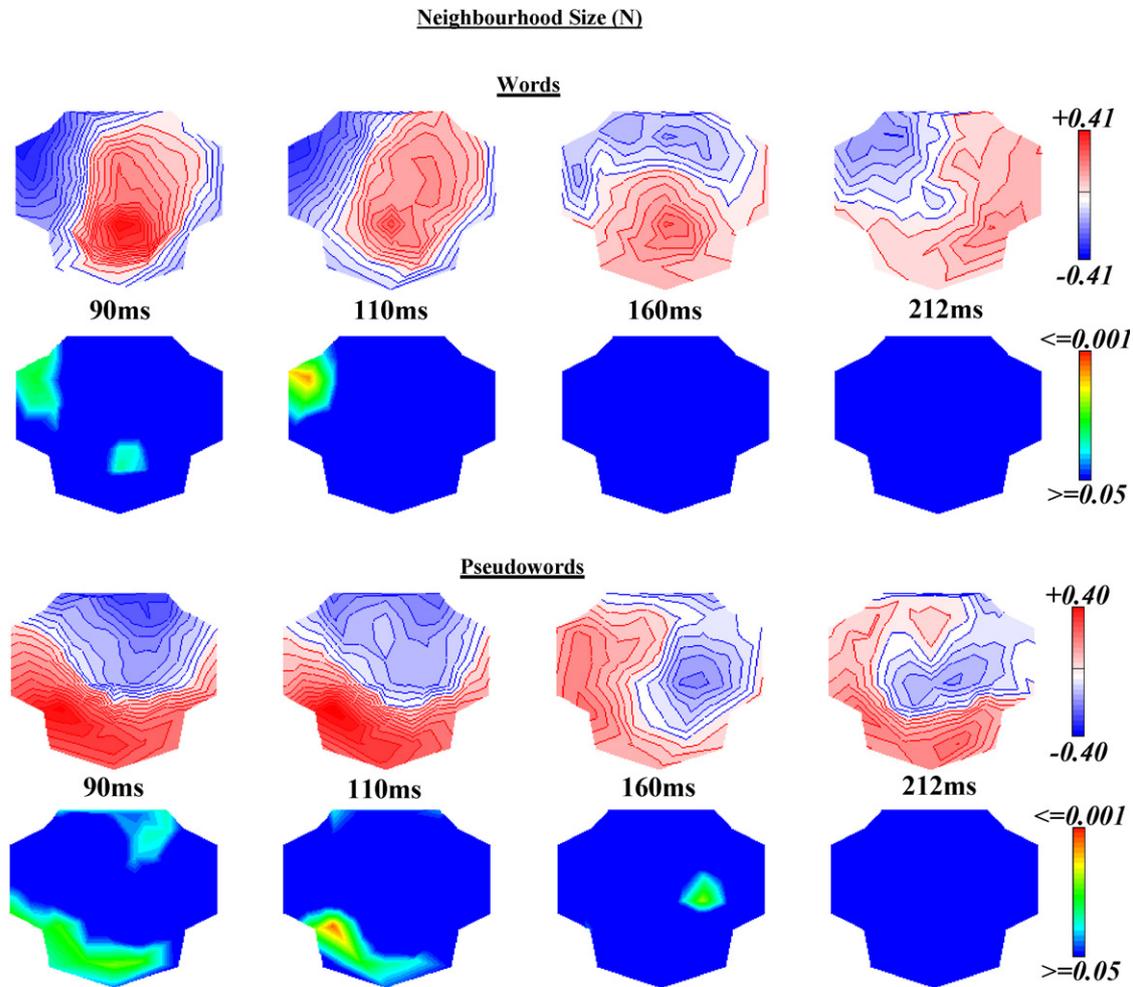
A previous study on ERP effects of  $N$  reported effects in lexical decision task in the time range of the N400 (350–550 ms), and these did not differ between words and pseudowords (Holcomb et al., 2002). In a separate experiment employing a semantic categorisation task, the same study found  $N$  effects in an earlier time range 150–300 ms. The difference between these and our findings can have several reasons. First, Holcomb et al. (2002) did not report results for a time window around 100 ms. Second, their earliest time window (150–300 ms) was 150 ms wide, which might therefore obscure short-lived effects such as those reported in this study. Note that other early effects of psycholinguistic variables in previous studies, as reviewed above, were usually reported for time points or time windows of only a few tens of milliseconds in duration (similar arguments have been made in the auditory domain, see Pulvermüller and Shtyrov, 2006). Third, as



**Fig. 2.** Grand-mean ERP topographies in the latency ranges of interest for words and pseudowords, respectively. Grey circles indicate approximate locations of electrodes used for topographical analyses.



**Fig. 3.** Grand-mean ERRC topographies for the variable 'Length', separately for words and pseudowords. The topographies of ERRCs are shown at the top of each panel, and the distribution of  $p$ -values from one-tailed  $t$ -tests against zero at the bottom.



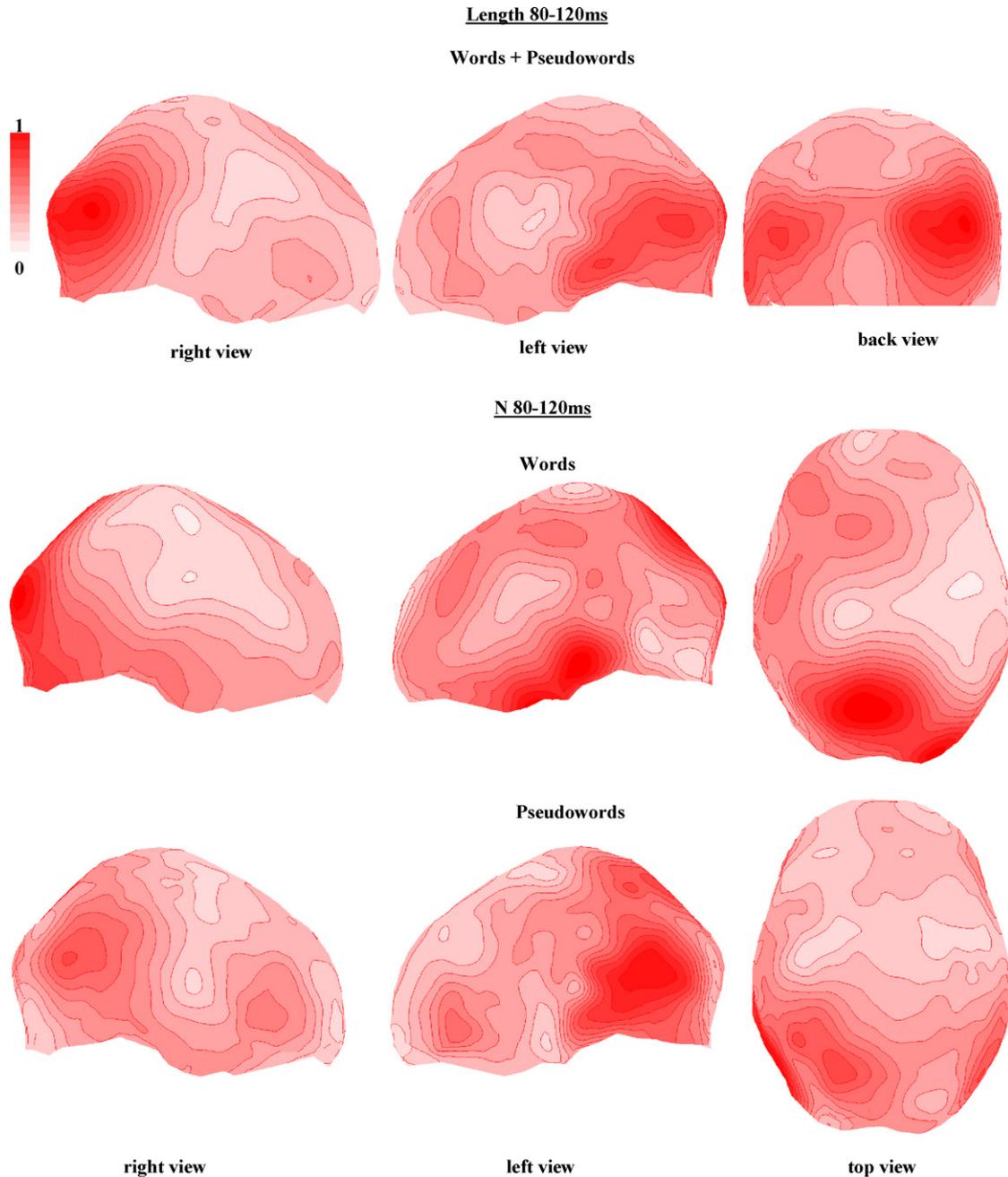
**Fig. 4.** Grand-mean ERRC topographies for orthographic neighbourhood size ( $N$ ), separately for words and pseudowords. The topographies of ERRCs are shown at the top of each panel, and the distribution of  $p$ -values from two-tailed  $t$ -tests at the bottom.

introduced at the outset, existing arguments make it likely that our multiple linear regression approach is more sensitive than factorial designs (e.g. Cohen, 1983). We therefore conclude that the results of this previous study and our present one do not contradict each other, but might reflect  $N$  effects at different stages of processing. Pykkänen et al. (2002) studied the effect of phonotactic probability (comparable to orthographic typicality) and phonological neighbourhood density (comparable to orthographic neighbourhood size) on MEG responses to visually presented words in the time ranges 140–220 ms (M170), 200–300 ms (M250), and 300–420 ms (M350). Only the latency of the M350 was reported to be sensitive to neighbourhood density. Results for a time window around 100 ms were not reported, which makes it difficult to compare our results with this study. Furthermore, no topographical analyses of these data were performed, but only amplitudes computed as root-mean-squares across a number of selected channels were analysed. If neighbourhood size did not affect the overall amplitude of the brain response, but its spatial pattern instead, this analysis would not be able to detect it.

It is interesting to note that the direct comparison between ERPs for words and pseudowords, as reported in our previous study revealed effects of lexicality around 160 ms, while effects of  $N$  in the present study differed between words and pseudowords earlier at around 100 ms (Hauk et al., 2006a). Similarly, a recent study found an interaction between lexicality and orthographic typicality around 160 ms, preceding a main effect of lexicality at

around 200 ms (Hauk et al., 2006a). This sequence of effects, i.e. an interaction between lexicality and psycholinguistic variables that quantify the ease or difficulty of form-based processing before a main effect of lexicality, speaks in favour of cascaded processing during word recognition. In this account, early stages of prelexical processing of written words overlap in time with later lexical access processes. The timing of access to lexical/semantic information (as reflected by a main effect of lexicality) is modulated by the structure of orthographic form representations (e.g. typicality, neighbourhood size). The difference in latencies for these effects can thus be explained by stimulus features which differed between the two studies. Hauk et al. (2006b) used pseudo-homophone pseudowords that were orthographically very similar to the word stimuli, e.g. in many cases differed by only one letter. Furthermore, all the real words had irregular spelling-sound correspondences (such as *yacht*). These two factors would be expected to slow word recognition, explaining the ERP latency differences (40–60 ms slower) and the longer reaction times (approximately 100 ms slower) compared to the present study and that of Hauk et al. (2006a).

One motivation for the present study was to explore the influence of two form-variables (word length and  $N$ ) that are strongly negatively correlated with each other both in the language as a whole, and in the present set of materials. Although multiple linear regression is able to determine their individual contributions to a given data set, the higher the correlation between two



**Fig. 5.** Minimum norm source estimates based on ERRCs for different psycholinguistic variables in the time window 80–120 ms. (Top row) Source estimate for 'Length', collapsed across words and pseudowords (from left to right: right/left/back view). (Middle) Source estimate for 'Neighbourhood size' ( $N$ ) for words (from left to right: right/left/top view). (Bottom) Source estimate for 'Neighbourhood size' for pseudowords (from left to right: right/left/top view). All distributions are normalised to their maximum value.

variables the more that noise will influence the results obtained (Bertero et al., 1988). One may therefore be concerned that the effects of both variables are driven by the same signal, and differences are caused by noise in the data. However, the effects of word length obtained in the present analysis replicated previous findings, so we can be confident that noise did not significantly affect the estimates for this variable. Furthermore, we found a lexicality effect on ERRCs for  $N$  but not for Length, demonstrating that the results for these two variables could not have been driven by the same underlying signal.

In addition to time course information, ERPs can provide at least coarse estimates of the spatial distribution of the observed effects. We should remember, however, that our source estimations were

based on 63-channel EEG recordings and an average head model which offer limited spatial resolution (well below that achieved with metabolic imaging such as fMRI) (e.g. Hauk, 2004). Source estimates of the early effects of Length were localised mainly to bilateral posterior brain areas for both words and pseudowords. Similar results have been obtained previously with EEG (Hauk et al., 2006a), MEG (Assadollahi and Pulvermüller, 2003), as well as with metabolic imaging (Mechelli et al., 2000). These localisations suggest that word length mainly affects basic visual processes, which do not differ between words and pseudowords and are not lateralised to the language-dominant hemisphere. For the variable  $N$ , in contrast, source estimates were left-lateralised, and the locations of peak activations differed between words and pseudo-

words. For words, activation was localised more anteriorly in the left temporal lobe compared to pseudowords. These topographical differences may be due to the fact that words activate their own memory trace in addition to their neighbours, while pseudowords are not associated with a memory trace. Activation in anterior temporal lobe may therefore reflect amodal lexico-semantic processing, e.g. competition and inhibition processes, related to the recognition of a letter string as part of the mental lexicon. fMRI localisations of activity associated with *N* are still sparse and inconsistent. Fiebach et al. (2007) reported effects related to *N* only in prefrontal brain areas, which was taken as evidence for a post-lexical origin of these effects. The pattern of our present results suggest that neighbourhood effects that occur at early stages of word processing are generated in temporal lobe areas previously associated with written word recognition, rather than in prefrontal regions associated with response selection. This is more consistent with the results of Binder et al. (2003), who found reduced brain activation for words with higher *N* in perisylvian areas, namely the left prefrontal gyrus, angular gyrus, and ventrolateral temporal areas.

In the behavioural literature, it is still debated whether effects of *N* reflect processes related to the retrieval of lexico-semantic information, such as competition between co-activated neighbours or global activation in the lexicon, or task-dependent strategies (e.g. Andrews, 1997; Balota et al., 2004; Grainger and Jacobs, 1996; Norris, 2006). In contrast to previous studies, our electrophysiological data showed effects of *N* already around 100 ms after word onset which distinguished between words and pseudowords, therefore arguing for an origin of *N* effects early in the word recognition process, rather than at a late strategy-dependent stage. Our source estimation results suggest the main generators of these effects in left perisylvian brain areas, providing further evidence that they reflect elementary processes in word recognition, rather than post-access phenomena. These results fit well to previous findings that other lexical and orthographic variables, such as word frequency and orthographic typicality, are reflected in the ERP within the first 200 ms after word presentation.

## Acknowledgments

We would like to thank Clare Dine for her help during data acquisition. We are also grateful to Maarten van Casteren for his advice on lexical databases. We would like to acknowledge the financial support of the U.K. Medical Research Council (U.1055.04.003.00001.01).

## References

- Andrews, S., 1989. Frequency and neighborhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15 (5), 802–814.
- Andrews, S., 1997. The effect of orthographic similarity on lexical retrieval: resolving neighborhood conflicts. *Psychonomic Bulletin and Review* 4 (4), 439–461.
- Assadollahi, R., Pulvermüller, F., 2003. Early influences of word length and frequency: a group study using MEG. *Neuroreport* 14 (8), 1183–1187.
- Baayen, R.H., Dijkstra, T., Schreuder, R., 1997. Singulars and plurals in Dutch: evidence for a parallel dual route model. *Journal of Memory and Language* 37, 94–117.
- Baayen, R.H., Piepenbrock, R., van Rijn, H., 1993. The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H., Yap, M., 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133 (2), 283–316.
- Bertero, M., De Mol, C., Pike, E.R., 1985. Linear inverse problems with discrete data. I. General formulation and singular system analysis. *Inverse Problems* 1 (4), 301–330.
- Bertero, M., De Mol, C., Pike, E.R., 1988. Linear inverse problems with discrete data. II. Stability and regularisation. *Inverse Problems* 4 (3), 573–594.
- Binder, J.R., McKiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., et al., 2003. Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience* 15 (3), 372–393.
- Cohen, J., 1983. The cost of dichotomization. *Applied Psychological Measurement* 7 (3), 249–253.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M.A., et al., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123 (Pt 2), 291–307.
- Coltheart, M., Davelaar, E., Jonasson, J.T., Besner, D., 1977. Access to the internal lexicon. In: Dornic, S. (Ed.), *Attention and Performance*, vol. 6. Academic Press, New York, pp. 535–555.
- Dambacher, M., Kiegl, R., Hofmann, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. *Brain Research* 1084 (1), 89–103.
- Davis, M.H., Meunier, F., Marslen-Wilson, W.D., 2004. Neural responses to morphological, syntactic, and semantic properties of single words: an fMRI study. *Brain & Language* 89 (3), 439–449.
- Dehaene, S., 1995. Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport* 6 (16), 2153–2157.
- Dien, J., Frishkoff, G.A., Cerbone, A., Tucker, D.M., 2003. Parametric analysis of event-related potentials in semantic comprehension: evidence for parallel brain mechanisms. *Brain Research Cognitive Brain Research* 15 (2), 137–153.
- Ellis, A.W., 2004. Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain & Language* 88 (3), 355–366.
- Fiebach, C.J., Ricker, B., Friederici, A.D., Jacobs, A.M., 2007. Inhibition and facilitation in visual word recognition: prefrontal contribution to the orthographic neighborhood size effect. *Neuroimage* 36 (3), 901–911.
- Fokas, A.S., Kurylev, Y., Marinakis, V., 2004. The unique determination of neuronal currents in the brain via magnetoencephalography. *Inverse Problems* 20, 1067–1082.
- Ford, M., Marslen-Wilson, W.D., Davis, M.H., 2003. Morphology and frequency: contrasting methodologies. In: Baayen, H., Schreuder, R. (Eds.), *Morphological Structure in Language Processing*. Mouton de Gruyter, Berlin/New York.
- Forster, K.I., Shen, D., 1996. No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (3), 696–713.
- Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage* 2 (2), 166–172.
- Fuchs, M., Wagner, M., Kohler, T., Wischmann, H.A., 1999. Linear and nonlinear current density reconstructions. *Journal of Clinical Neurophysiology* 16 (3), 267–295.
- Gernsbacher, M.A., 1984. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General* 113 (2), 256–281.
- Golub, G.H., van Loan, C.F., 1996. *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, Baltimore/London.
- Grainger, J., Jacobs, A.M., 1996. Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review* 103 (3), 518–565.
- Graves, W.W., Grabowski, T.J., Mehta, S., Gordon, J.K., 2007. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience* 19 (4), 617–631.
- Gruber, T., Trujillo-Barreto, N.J., Giabbiconi, C.M., Valdes-Sosa, P.A., Muller, M.M., 2006. Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *Neuroimage* 29 (3), 888–900.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1984. Interpreting Measured Magnetic Fields of the Brain: Minimum Norm Estimates of Current Distributions. Helsinki University of Technology, Technical Report TTK-F-A559.
- Hauk, O., 2004. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage* 21 (4), 1612–1621.
- Hauk, O., Davis, M.H., Ford, M., Pulvermüller, F., Marslen-Wilson, W.D., 2006a. The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30 (4), 1383–1400.
- Hauk, O., Keil, A., Elbert, T., Muller, M.M., 2002. Comparison of data transformation procedures to enhance topographical accuracy in time-series analysis of the human EEG. *Journal of Neuroscience Methods* 113 (2), 111–122.
- Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., Rogers, T.T., 2006b. [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 msec ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience* 18 (5), 818–832.
- Hauk, O., Pulvermüller, F., 2004a. Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology* 115 (5), 1090–1103.
- Hauk, O., Pulvermüller, F., 2004b. Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping* 21 (3), 191–201.
- Hauk, O., Shtyrov, Y., Pulvermüller, F., 2003. Localising the average or averaging the localisations? It does not matter for the early left-anterior Mismatch Negativity! Abstract presented at the 9th International Conference on Functional Mapping of the Human Brain in New York, USA. *Neuroimage* 19 (2) supplement 1 poster number 1297.
- Hinojosa, J.A., Martin-Loeches, M., Munoz, F., Casado, P., Pozo, M.A., 2004. Electrophysiological evidence of automatic early semantic processing. *Brain & Language* 88 (1), 39–46.
- Holcomb, P.J., Grainger, J., O'Rourke, T., 2002. An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience* 14 (6), 938–950.

- Howell, D.C., 2001. *Statistical Methods for Psychology*. Duxbury.
- Ilmoniemi, R.J., 1993. Models of source currents in the brain. *Brain Topography* 5 (4), 331–336.
- Kissler, J., Herbert, C., Winkler, I., Junghöfer, M., 2008. Emotion and attention in visual word processing—an ERP study. *Biological Psychology*, in press.
- Landauer, T.K., Dumais, S.T., 1997. A solution to Plato's problem: the latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104, 211–240.
- Martin-Loeches, M., Hinojosa, J.A., Gomez-Jarabo, G., Rubia, F.J., 1999. The recognition potential: an ERP index of lexical access. *Brain & Language* 70 (3), 364–384.
- Mechelli, A., Humphreys, G.W., Mayall, K., Olson, A., Price, C.J., 2000. Differential effects of word length and visual contrast in the fusiform and lingual gyri during reading. *Proceedings of the Royal Society of London B: Biological Sciences* 267 (1455), 1909–1913.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave De Peralta, R., 2004. EEG source imaging. *Clinical Neurophysiology* 115 (10), 2195–2222.
- Norris, D., 2006. The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review* 113 (2), 327–357.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- Parks, R., Ray, J., Bland, S., 1998. Wordsmyth English Dictionary-Thesaurus. <http://www.wordsmyth.net/>.
- Penolazzi, B., Hauk, O., Pulvermüller, F., 2007. Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology* 74 (3), 374–388.
- Perea, M., Carreiras, M., Grainger, J., 2004. Blocking by word frequency and neighborhood density in visual word recognition: a task-specific response criteria account. *Memory & Cognition* 32 (7), 1090–1102.
- Pulvermüller, F., Lutzenberger, W., Birbaumer, N., 1995. Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology* 94 (5), 357–370.
- Pulvermüller, F., Mohr, B., Schleicher, H., 1999. Semantic or lexico-syntactic factors: what determines word-class specific activity in the human brain? *Neuroscience Letters* 275 (2), 81–84.
- Pulvermüller, F., Shtyrov, Y., 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology* 79 (1), 49–71.
- Pylkkänen, L., Stringfellow, A., Marantz, A., 2002. Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain & Language* 81 (1–3), 666–678.
- Rudell, A.P., 1991. The recognition potential contrasted with the P300. *International Journal of Neuroscience* 60 (1–2), 85–111.
- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology* 32 (1), 11–22.
- Scherg, M., 1992. Functional imaging and localization of electromagnetic brain activity. *Brain Topography* 5 (2), 103–111.
- Scott, G.G., O'Donnell, P.J., Leuthold, H., Sereno, S.C., 2008. Early emotion word processing: evidence from event-related potentials. *Biological Psychology*, in press.
- Sears, C.R., Hino, Y., Lupker, S.J., 1995. Neighborhood size and neighborhood frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21 (4), 876–900.
- Sears, C.R., Lupker, S.J., Hino, Y., 1999. Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: a test of the multiple read-out model. *Perception & Psychophysics* 61 (8), 1537–1554.
- Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9 (10), 2195–2200.
- Siakaluk, P.D., Sears, C.R., Lupker, S.J., 2002. Orthographic neighborhood effects in lexical decision: the effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance* 28 (3), 661–681.
- Skrandies, W., 1998. Evoked potential correlates of semantic meaning—a brain mapping study. *Brain Research Cognitive Brain Research* 6 (3), 173–183.
- Weekes, B.S., 1997. Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 50 (2), 439–456.
- Whaley, C., 1978. Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17, 143–154.