



ERP signal analysis with temporal resolution using a time window bank

Annika Nijveld¹, Louis ten Bosch^{1,2}, Mirjam Ernestus¹

¹CLS, Radboud University, Nijmegen, the Netherlands

²CLST, Radboud University; Donders Institute, Nijmegen, the Netherlands

a.nijveld@let.ru.nl, l.tenbosch@let.ru.nl, m.ernestus@let.ru.nl

Abstract

In order to study the cognitive processes underlying speech comprehension, neuro-physiological measures (e.g., EEG and MEG), or behavioural measures (e.g., reaction times and response accuracy) can be applied. Compared to behavioural measures, EEG signals can provide a more fine-grained and complementary view of the processes that take place during the unfolding of an auditory stimulus.

EEG signals are often analysed after having chosen specific time windows, which are usually based on the temporal structure of ERP components expected to be sensitive to the experimental manipulation. However, as the timing of ERP components may vary between experiments, trials, and participants, such a-priori defined analysis time windows may significantly hamper the exploratory power of the analysis of components of interest. In this paper, we explore a wide-window analysis method applied to EEG signals collected in an auditory repetition priming experiment.

This approach is based on a bank of temporal filters arranged along the time axis in combination with linear mixed effects modelling. Crucially, it permits a temporal decomposition of effects in a single comprehensive statistical model which captures the entire EEG trace.

Index Terms: EEG, ERP analyses, speech comprehension, time window analysis, linear mixed effects models

1. Introduction

Neurological measurements such as EEG signals play an important role in present-day psycholinguistic research. Especially in combination with behavioural data, EEG signals can provide a fine-grained and complementary view of the cognitive processes taking place during the unfolding of an auditory stimulus.

EEG differs from behavioural measures in at least two important ways. First, it registers processes much earlier in time (namely, directly and on-line compared to behavioural overt effects hundreds of milliseconds later). This is advantageous given that effects of cognitive processes may manifest long before a participant's overt response. Event Related Potential (ERP) components distinguish earlier from later processing stages during the presentation of a stimulus, which may be cancelled out or minimally detectable afterwards in behavioural measures. This is especially relevant when a stimulus itself unfolds over time, such as with auditory input. Second, where behavioural responses are the outcome of an accumulation of processes, including the motor processes required for making an overt response, EEG signals may tap into a single processing stage. As such, the EEG method is considered to offer a closer estimate of processes related to perception. Moreover, specific ERP components have been functionally associated with distinct stages of spoken word recognition, which makes it possi-

ble to draw conclusions about which stages of processing are affected by experimental manipulations. For instance, the N400 component has been used to probe semantic match [1] as well as lexical processing, going from activation of a set of word candidates to the selection of the target word [2].

The motivation for our original study is the use of EEG signals to unravel subtle types of cognitive processing differences that take place during the auditory presentation of words. In addition, we used EEG to probe processing advantages of a word token after having heard another token of the same word recently, and differences therein depending on the degree of acoustic similarity between these tokens ('exemplar' or 'specificity' effects, previously established in ERPs by [3, 4]). Specifically, we investigated whether such processing advantages are task-dependent, as behavioral work suggests [5].

In this paper, we will investigate the role of the position of the EEG analysis window during the unfolding of target words using the data of our auditory long-term priming experiment. Each experimental word occurred as prime (first occurrence) and as target (second occurrence), and prime and target could be uttered by the same ('matching') or a different ('mismatching') speaker (see Table 2). Listeners either had to perform old-new judgments or semantic judgments. Specifically, we hypothesized that prime-target matching is more important in the old-new judgment task (in which participants need to specifically use the memory trace of the prime) than in the semantic classification task.

In classical EEG studies, researchers often focused on specific time windows for the EEG analysis. These windows are chosen based on the significance of effects in specific windows reported in the literature. However, as the profile of some effects and/or components may vary in time between experimental paradigms, participants, and trials (as discussed by [6]), this window-based approach may not be optimal in all cases. This problem is amplified if the literature provides no clear time window precedent for a given effect. One solution to this problem is cluster-based permutation tests [7], but this approach is not applicable to data involving different participant groups, as in our experiment. Other types of EEG analyses are based on wavelets [8, 9], which allow the simultaneous use of different *time-frequency scales*, given the choice of an analysis window.

In this paper, we propose a method that can be applied to different participant groups while specifically circumventing a-priori time-window dependency. This is done by introducing an interaction by means of a window 'bank' specified along the time axis as a predictor in interactions, and by investigating the significance of this interaction as a function of time.

The next sections will present the data, methods and results, after which a discussion and conclusion section follows.

2. Experimental data

2.1. Participants

The dataset contained data from 46 right-handed Dutch native participants, 22 randomly assigned to the old-new judgment task (6 male, 18-30 years old, mean: 21 years old), and 24 to the semantic classification task (3 male, 19-30 years old, mean: 22 years old). All were highly educated, were paid for their participation, reported good hearing, and had no known neurological disorders. They gave their written informed consent.

2.2. Materials

The experimental words were 64 inanimate Dutch bisyllabic nouns (e.g., *gieter* ‘watering can’) with an average log-transformed frequency of occurrence of 2.2 ($SD = 0.8$, counts obtained from [10]). The experiment also contained 64 animate filler words that were matched to the experimental words in frequency of occurrence (e.g., *kikker* ‘frog’). We recorded primes with a male and female native speaker of Dutch, and targets and fillers with the male speaker. The loudness of all stimuli was set to 70 dB.

2.3. Experimental set-up

The experiment consisted of two parts, each consisting of a familiarization and a test phase (see also Table 1). During each familiarization phase, we presented 32 experimental words as primes. During each test phase, we presented the 32 experimental words from the familiarization phase as targets, in addition to 32 fillers. The division of the experiment into two parts served to keep primes and targets close together, as previous research has shown that a separation of primes and targets by too many intervening trials (i.e., > 100 trials) diminishes the chances of finding exemplar effects [11].

Half of the primes were presented in the voice of the male speaker, and half in the voice of the female speaker. All targets were presented in the voice of the male speaker (see also Table 2). Primes and targets were always different tokens. To not stand out from the targets, fillers (occurring in the test phases alongside the targets) were also all uttered by the male speaker.

Each participant received a counterbalanced list of trials in which no more than three consecutive primes in the same speaker voice occurred, and no more than three consecutive targets or fillers occurred. All primes and targets were separated by no more than 100 trials.

2.4. Procedure

We tested participants individually in a sound-attenuating booth, and presented auditory stimuli via closed headphones at a comfortable listening level using Presentation software (Version 16.4, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

In the familiarization phases, all participants engaged in the same task, which focused their attention to perceptual aspects of the primes (beneficial for observing exemplar effects on the targets [12]). For the test phases, participants were divided among two groups. The participants in one group were instructed to indicate as quickly and accurately as possible whether the word they heard had occurred previously in the experiment or not (‘old’-responses with the dominant hand). Participants in the other group were instructed to indicate as quickly and accurately as possible whether the word referred to a living or a non-living object (‘non-living’-responses with the dominant hand).

Table 1: Overview of tasks and stimuli used in the familiarization and test phases of the two parts of the experiment. Old-new denotes the old-new judgment task, *sem. class.* the semantic classification task.

Part 1			
Fam. phase 1		Test phase 1	
task: loudness judgment		task: old-new <i>or</i> sem. class.	
32 primes	<i>tafel</i> ‘table’	32 targets	<i>tafel</i> ‘table’
		32 fillers	<i>oma</i> ‘grandmother’
Part 2			
Fam. phase 2		Test phase 2	
task: loudness judgment		task: old-new <i>or</i> sem. class.	
32 primes	<i>anker</i> ‘anchor’	32 targets	<i>anker</i> ‘anchor’
		32 fillers	<i>vlinder</i> ‘butterfly’

Table 2: Experimental conditions and speaker gender of primes and targets in the experiment.

Condition	Prime	Target
Speaker match	male	male
Speaker mismatch	female	male

2.5. EEG recording and preprocessing

We collected EEG data from 64 Ag-Ag CI active electrodes with the 10-20 actiCAP system (Brain Products GmbH, Munich; 8 midline and 50 lateral electrodes). We kept electrode impedances below 50 k Ω . The EEG was recorded continuously with a 0.02 - 100 Hz band-pass filter, and resampled with a 500 Hz sampling frequency. Using FieldTrip [13], data were offline re-referenced to the mastoids, and band-filtered with 35 Hz cut-off. Noisy trials were manually rejected, and ocular artefacts were removed via ICA.

2.6. Analyses

The conventional time window-based analysis, using data based on an a-priori time window (data[time window]), can be represented by a linear mixed effects model of the following type (model-I):

```
model-I =
  lmer(amplitude ~ task * match
    + sample + (1 | participant) +
    (1 | word),
    data = data[time window])
```

in which *task* and *match* (both 2-level categorical predictors) denote the participants’ task, and the prime-target matching with respect to speaker voice, respectively. The dependent variable (*amplitude*) refers to the *mean* over time of each individual EEG trace; the mean is taken over the time window specified. The predictor *sample* denotes the sample id (its range depends on the time window used). *Participant* and *word* serve as random factors. Here, the slopes are not shown for brevity: their inclusion was taken into account depending on the dataset. The dataset is specified as data[time window] to make clear that it depends on the specific window chosen (e.g., 100-300 ms, 300-400 ms, 400-800 ms from word onset).

3. Results

In all analyses, models were applied in two ways: to both all trials, and to the subset of trials that received correct responses. These results were highly similar; here, we only report the models applied to the trials that received correct responses. In all reported models, outlier data (i.e., data points with standardized residuals exceeding 2.5 standard deviation units) were removed, after which we refitted the models.

3.1. Analysis with predefined time windows

Previous studies showed exemplar effects in ERPs on different components: the N100/P200 (associated with early acoustic processing) [14], P350 (associated with lexical identification) [15, 14], and N400 (associated with lexical selection) [3, 14]. In our original study, we therefore tested for exemplar effects in type model-I analyses on three pre-defined time windows and scalp topographies corresponding to these components: 100-300 over posterior electrodes, 300-400 over posterior electrodes, and 400-800 over all electrodes. These analyses (reproduced in Table 3) yielded the clearest task-driven exemplar effects in the 100-300 window: the *task * match* interaction was highly robust in that window. In the 300-400 window, we also observed a statistically significant *task * match* interaction, albeit less robust. In the 400-800 window, the interaction was not statistically significant.

Table 3: Results of model-I for three windows (ms from onset). For the sake of clarity, only the fixed effects of interest are shown. The intercept represent the old-new judgment task and speaker mismatch, 'Sem. class. task' stands for the semantic classification task. A dash indicates no significance.

	100-300		300-400		400-800	
Fixed effects	$\hat{\beta}$	<i>t</i>	$\hat{\beta}$	<i>t</i>	$\hat{\beta}$	<i>t</i>
Intercept	-1.07	-3.6	-1.93	-5.4	-5.64	-17.4
Sem. class. task	0.35	0.9	0.60	1.3	-	-
Match	-0.36	-4.8	-0.17	-1.9	-	-
Sem. class. task * match	0.56	5.8	0.28	2.4	-	-

3.2. Analysis using a time window bank

The time window bank integrates the results of different time windows. We therefore keep the EEG amplitudes, without averaging them as is done for the model-I analyses. The dependent variable in model-II type regressions is a stretch of an EEG trace, instead of the averaged EEG amplitude within a window. The advantage is that amplitudes changing over time are treated in one model that takes all effects into account, without information loss due to averaging. It also improves approaches based on individual *t*-tests per sample as were frequently used in the more classical EEG studies, since the comprehensive model does not require correction for multiple comparisons.

Figure 2 shows the result of model-II. The thin red solid line, in combination with the transparent confidence interval, reflects the significance levels of the 3-way interaction *task * match * tw*, as provided by model-II, for electrode Cz. The *t* values positioned are at the center point (ms) of the corresponding *tw*-column.

The outcome of model-II result can be validated through comparison with the outcome of model-I as ran on the same electrode (Cz). To validate model-II in this way for a time point *t*, model-I can be applied on a time window centered around

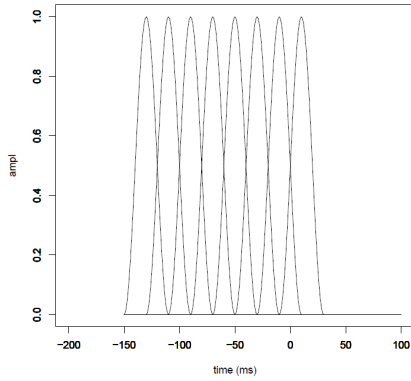


Figure 1: Example of a time window bank. The figure shows the first 8 time windows of a bank used as predictor in a three-way interaction with task and match. Time is displayed along the horizontal axis. Each single plot corresponds to a column in the matrix *tw* in regression model model-II. $T = 0$ refers to word onset. Here, time window spacing is 20 ms.

The drawback of a model-I approach is the a-priori choice of a time analysis window. In addition, this approach may lead to issues concerning multiple comparisons in several time windows. To circumvent this, we take model-I as starting point but extend it to include an additional predictor *time window* (denoted *tw*). This predictor is positioned in interaction with our 2-way interaction of interest (*task * match*). This inclusion leads to the extended type of regression model (model-II) (using R-syntax and nomenclature):

```
model-II =
  lmer(amplitude ~ task * match * tw
    + ave + sample + (1 | participant)
    + (1 | word),
    data = data[wide time window])
```

The predictor *tw* must be constructed in such a way that multiple different applications of model-I on different time windows of interest can be replaced by a single coherent application of model-II on a larger time window, that covers all windows of interest.

In model-II, *tw* consists of a rectangular matrix. Its number of rows matches the length of the EEG signals (which is the length of the chosen overall time window, here chosen to be -199 ms to $+800$ ms relative to word onset). Its number of columns equals the number of time windows used. Each column represents one time window, consisting of half a period of a squared cosine with lowest value 0 and peak value equal to 1 (see Figure 1 for an example showing a few of these time windows aligned). The whole matrix *tw* consists of systematically shifted versions of a column, with the time window peak location regularly shifted across columns, determined by the window spacing (20 ms in Figure 1), between -199 and 800 ms after onset.

The inclusion of *ave* in model-II allows this model to attribute deviations from the average EEG trace to specific predictors, and in particular, to investigate the significance of the interaction of *task* and *match* (that we are after) as a function of time.

t . The outcomes of model-I are shown as blue symbols in Figure 2. For example, the blue symbols at $t = 0$ are determined by applying model-I using the time windows with lengths specified in the legend. In this way, the outcome of model-II can be validated for each t between -200ms to 800ms.

The figure shows that model-II does fairly well in predicting the outcomes of model-I for various time points t . The figure might suggest that model-I can safely be applied *repeatedly* using a shifting window. That would be incorrect, however, as statistical issues with multiple comparisons should be avoided. The blue symbols only serve to validate the outcome of model-II across the entire time window. The blue symbols also show that the t -values depend slightly on the size of the window chosen. Most importantly, however, the figure indicates that the model-II regression can be used to explain the results obtained via the a-priori window approach (as used in the models in Table 3).

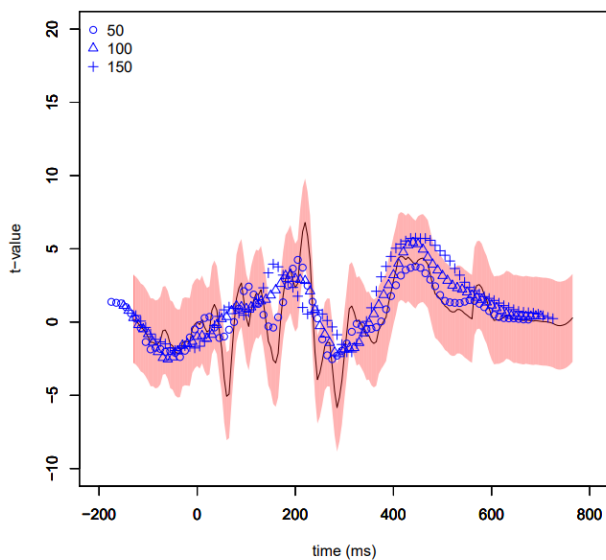


Figure 2: This plot combines the significance levels of the interaction $task * match$ for model-I (separate sliding windows; blue symbols) and model-II (by the time window bank; solid line and transparent confidence interval). The legend specifies the window lengths in the sliding window approaches.

4. Discussion and conclusion

In this paper, we focused on the role of the position of the EEG analysis window during the unfolding of target words in an auditory priming study, and we proposed a time bank-based statistical method to enable a statistical model to perform one coherent time course analysis within a large time window. To that end, we compared two types of linear mixed effect regression models: model-I based on separate individual time windows, and model-II using an integrative bank of time windows.

Figure 2 shows a good agreement (correlation ≈ 0.86) between the t values obtained using different windows from model-I, and the t values obtained from the single application of model-II on the larger time window. The model-II results are in agreement with the observation that the model-I results in the second time window (300-400 ms) were less robust than in the first window (100-300 ms). In combination with the confi-

dence intervals, model-II does more: it also explains why in the model-I approach, the time window 400-800 ms leads to weak or absent significance levels for the interaction of interest. The effect of interest may be highly significant in certain narrow time intervals (here, seemingly around 500 ms), but may disappear when larger time windows are applied, in which critical but local patterns get lost.

Since the t values specify the significance levels of the $task * match$ interaction (under the usual assumption that the residuals obey a t distribution), the plots show that significance levels can reliably be estimated using a regularly spaced time bank. As a consequence, model-II provides, in combination with the confidence intervals, a more consistent, statistically better defensible analysis of the time course. To avoid fishing, a model-II analysis should not be used for steering a window-based model-I analysis.

The construction of the time windows in the bank is relevant. Here, a squared cosine was chosen, since this allows the sum of neighbouring shifted versions to equal unity for the entire window -200-800 ms.

It is interesting to note that a similar analysis using the `bam()` function in R with *trial* as a smooth yielded virtually the same t values. With `gam()/bam()` functions, warped versions of predictors can be included into the model, which allows for local compression and stretching of, e.g., the time axis. In the current model-II analysis, however, the statistical trends over time can be made explicit per time interval. This might be relevant for the close examination of early and late processes: previous auditory and cross-modal studies established ERP exemplar effects in early as well as late time windows relative to word onset. These effects were related to the N100/P200 complex (around 100-300 ms), P350 (around 300-400 ms), and N400 components (around 400-800 ms). Often, effects of interest arising on different ERP components change in polarity over the course of the trial. The time point of components and polarity changes can often not be predicted exactly; analyses based on a-priori chosen analysis time windows may therefore not adequately capture them.

Importantly, the current analysis does not make any a-priori assumptions about polarity changes of effects over time. Figure 2 shows polarity patterns, but they relate to the t values, rather than to the underlying EEG traces. Patterns we observed in the 300-400 ms window may have the opposite polarity from those in the 100-300 ms time window, probably due to the upcoming N400 component, which is among other things related to *lexical* processing. The change of polarity of exemplar effects over time as expressed on different components is topic for future research.

This paper contributes to the study of the effect of the time window in EEG analyses. The EEG research field is rather complex, with many effects and factors playing a role simultaneously and/or consecutively in a rich data space. Chopping the problem into smaller sub-problems is therefore essential for a better grip on the underlying processes. We restricted the current analysis to the variation of effects over time, while leaving aside the variation of effects over electrode sites (which, in turn, is likely to interact with effects' variation in time).

Another topic for future investigation is therefore the use of the electrode sites. In this paper, we took Cz in model-II as representing the different multiple electrode sets used in model-I. The inclusion of multiple electrode sites often leads to quite complex statistical models, which do not always converge. The choice in this paper was taken as a reasonable and defensible starting point for further investigation.

5. References

- [1] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP)," *Annual Review of Psychology*, vol. 62, pp. 621–647, 2011.
- [2] A. S. Desroches, R. L. Newman, and M. F. Joanisse, "Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity," *Journal of Cognitive Neuroscience*, vol. 21, no. 10, pp. 1893–1906, 2009.
- [3] S. Dufour, D. Bolger, S. Massol, P. J. Holcomb, and J. Grainger, "On the locus of talker-specificity effects in spoken word recognition: an ERP study with dichotic priming," *Language, Cognition and Neuroscience*, vol. 32, no. 10, pp. 1273–1289, 2017.
- [4] S. Campeanu, F. I. M. Craik, K. C. Backer, and C. Alain, "Voice reinstatement modulates neural indices of continuous word recognition," *Neuropsychologia*, vol. 62, pp. 233–244, 2014.
- [5] P. A. Luce and E. A. Lyons, "Specificity of memory representations for spoken words," *Memory and Cognition*, vol. 26, no. 4, pp. 708–715, 1998.
- [6] S. J. Luck, "Ten simple rules for designing ERP Experiments," in *Event-related potentials - A methods handbook*, T. C. Handy, Ed. Cambridge, MA: MIT press, 2005, pp. 17–32.
- [7] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.
- [8] O. A. Rosso, M. T. Martin, A. Figliola, K. Keller, and A. Plastino, "EEG analysis using wavelet-based information tools," *Journal of Neuroscience Methods*, vol. 153, pp. 163–182, 2006.
- [9] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure*, vol. 26, pp. 56–64, 2015.
- [10] E. Keuleers, M. Brysbaert, and B. New, "SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles," *Behavior Research Methods*, vol. 42, no. 3, pp. 643–650, 2010.
- [11] I. Hanique, E. Aalders, and M. Ernestus, "How robust are exemplar effects in word comprehension?" *The Mental Lexicon*, vol. 8, no. 3, pp. 269–294, 2013.
- [12] S. D. Goldinger, "Words and voices: episodic traces in spoken word identification and recognition memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 5, pp. 1166–1183, 1996.
- [13] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Computational Intelligence and Neuroscience*, vol. 2011, 2011.
- [14] U. Schild, A. B. Becker, and C. K. Friedrich, "Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition," *Frontiers in Psychology*, vol. 5, pp. 1–12, 2014.
- [15] C. K. Friedrich, S. A. Kotz, A. D. Friederici, and K. Alter, "Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence," *Cognitive Brain Research*, vol. 20, no. 2, pp. 300–308, 2004.