Electrophysiological Methods

Joost Rommers and Kara D. Federmeier

In Press, In De Groot, A. M. B., & Hagoort, P. (Eds.), Research Methods in Psycholinguistics: A Practical Guide. Hoboken, NJ: Wiley Blackwell.

Abstract

Recordings of electrical brain activity allow researchers to track multiple cognitive subprocesses with high temporal resolution. This chapter discusses how the electroencephalogram (EEG) is generated and recorded, and how it is analyzed, including filtering, artifact rejection, and statistical testing. It shows how electrophysiological methods have been used to study language, including discussion of aspects of experimental design, stimuli, and tasks, illustrated with a concrete example study. The chapter ends with some advantages and disadvantages of electrophysiological methods and current developments in their use. It is concluded that the noninvasive measurement of electrical brain activity generates some of the most direct evidence regarding the processes underlying language comprehension, production, and acquisition in the brain. The methods are likely to continue to provide important new insights that challenge our views of cognition and brain functioning.

Keywords: electrical brain activity, event-related brain potentials, ERPs, EEG, methods, language comprehension, language production, language acquisition

Language processing is multifaceted and unfolds rapidly, necessitating methods that can reveal the operation of multiple cognitive subprocesses with high temporal resolution. One such method, which has played a critical role in developing our understanding of language over the last several decades, is the recording of electrical brain activity through the electroencephalogram (EEG). This chapter discusses how electrophysiological methods have been used to study language, their advantages and disadvantages, and current developments in their use.

Underlying Assumptions and Rationale

The human EEG was discovered in the 1920s, when Hans Berger recorded and amplified electrical activity from the surface of a patient with head injury (Millett, 2001). Among other phenomena, Berger observed the alpha wave, an oscillation of around 10 cycles per second that was most prominent when the eyes were closed. Berger's findings were initially met with skepticism, but were eventually replicated. Now, countless studies have used EEG to investigate virtually all aspects of cognition, including perception, action, attention, memory, and language.

The EEG signal is a direct, continuous measure of brain activity. One of its primary strengths lies in its temporal resolution, which is on the order of milliseconds. This temporal resolution makes it one of the methods of choice for answering "when" questions in psycholinguistics. How long after a word is encountered is it integrated with its context? Is syntactic information retrieved before phonological information during word production? Do language-specific phoneme categories impact early or late speech perception processes?

Most EEG-based studies on language processing have relied on the derivation of event-related potentials (ERPs) from the ongoing data. ERPs are created by extracting from the continuous EEG data the brain responses time-locked to an event of interest, such as the onset of a stimulus or response. Typically, epochs from multiple instances of the same or similar events are aligned and averaged together point-by-point, so that random fluctuations in the EEG signal will tend to cancel one another, revealing the stable event-related voltage fluctuations. As shown in Figure 1, plotting the averaged voltage changes over time – the ERP – reveals a pattern of positive and negative deflections that can be linked to specific neural systems and functional processes; waveform features with well-established links are often referred to as “ERP components”. The timing and amplitude of ERP components have been shown to be sensitive indices of changes in specific cognitive processes related to stimulus perception and evaluation, attentional allocation, memory encoding and retrieval, response selection, motor preparation, and error- and reward- related processing, among others (see Luck & Kappenman, 2011).

< Insert Figure 1 about here >

Most ERP components are labeled according to their polarity and the (typical) latency or ordinal position of the peak. For example, the P200 (or P2) is a positive peak that occurs around 200 ms after onset of a visual stimulus. Many components also have a characteristic scalp distribution that helps identify them. The ERP component that has been used most in language research is probably the N400, a centro-parietally distributed negative-going waveform feature that peaks around 400 ms after the onset of potentially meaningful stimuli. The N400 was initially discovered as a response to unexpected words in sentences, being larger in amplitude to “dog” than “sugar” at the end of a sentence such as “I drink my coffee with cream and…” (Kutas & Hillyard, 1980). However, the N400 is not an anomaly detector. Subsequent studies have established that the N400 is part of the normal response to content words in all modalities, as well as to pictures and other meaningful stimuli, with its amplitude being attenuated as a function of the contextual support for the stimulus (for review, see Kutas & Federmeier, 2000, 2011). For instance, the N400 amplitude to a word in a sentence is inversely related to the word's cloze probability, operationalized as the proportion of participants who would provide that word as a continuation when given the sentence fragment in an off-line task. The N400 also decreases with repetition, word position in a congruent sentence, semantic relatedness to a preceding word in a list, and even semantic relatedness to expected but not actually presented words. Across all of these manipulations, the latency of the N400 is remarkably stable, unlike some other components (such as the P300) whose timing depends on various aspects of the experimental manipulation.

A second component that has often been used in language research is the P600, a longer-lasting positivity with a less consistent timing that does not always exhibit a clear peak. It was initially reported as a response to grammatical violations such as "throw" in "The spoilt child throw the toys on the floor" (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), opening up the possibility of tracking grammatical processing with ERPs and suggesting the possibility of a neural dissociation between semantics (associated with N400 effects, as just discussed) and syntax. However, later studies reported similar effects to spelling errors (Münte et al., 1998) and semantic reversal anomalies such as "For breakfast the eggs would only eat" (Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). These findings shifted the view of the P600 towards revision or repair processes, although several different interpretations currently exist (e.g., Brouwer, Fitz, & Hoeks, 2012; Coulson, King, & Kutas, 1998; Kolk & Chwilla, 2007; Kuperberg, 2007), including those that link the P600 to domain-general responses like the P300. Another component that has been linked to syntactic processing is the Left Anterior Negativity

(LAN), occurring at around 300-500 ms with a left frontal distribution (Osterhout & Holcomb, 1992). It has been reported in response to agreement errors and has been linked to morphosyntactic processing (Friederici, 1995) but also to working memory (Kluender & Kutas, 1993). Recently, however, it has been suggested that at least some apparent LAN effects could also arise from component overlap between an N400 and the onset of a right-lateralized P600 (Tanner, 2015).

Although some components tend to reliably be associated with peaks in the ERP, an individual waveform typically does not allow the researcher to draw conclusions about cognitive processes. Instead, as with most methods, the focus is on differences between conditions, or “ERP effects”. An ERP effect is a modulation of an ERP component, or just the difference between two conditions, which, in well-designed studies, isolates particular subprocesses of interest.

The focus of the sentence processing literature on the N400, P600 and LAN certainly does not mean that these are the only important components for studies of language. In fact, it should be stressed that language manipulations routinely elicit ERP effects that are not specific to language, because so many cognitive functions come together when reading, listening, or speaking. Moreover, among some of the most elegant studies using ERPs to answer questions in language processing are those that have made use of components that were originally characterized in very different contexts. For example, the Lateralized Readiness Potential (LRP), a component associated with response selection, has been used to study timing questions in language production (van Turennout, Hagoort, & Brown, 1997). Furthermore, the Mismatch Negativity (MMN), a component associated with auditory sensory memory, has been used to study phonological processing (Dehaene-Lambertz, 1997; Näätänen et al., 1997). This emphasizes the utility of being aware of the full toolbox of electrophysiological responses that could potentially be harnessed to do psycholinguistics (for a thorough overview, see Luck & Kappenman, 2011).

Apparatus

The voltage changes in the EEG are a direct, instantaneous measure of neural activity. The signal is thought to arise primarily from post-synaptic potentials produced by large populations of cortical pyramidal neurons that fire in synchrony. Pyramidal cells are the likely main contributor to the EEG signal because they occur in layers close to the scalp and are oriented in a common direction, which allows the activity from multiple neurons to summate rather than cancel out. The relatively slower post-synaptic potentials are a more probable source of the signal than action potentials, because action potentials are of short duration and therefore less likely to occur in synchrony (see Nunez & Srinivasan, 2006).

Non-invasive recordings of these potentials are possible through the use of silver-silver chloride or tin electrodes affixed to or held near the face and scalp (via, for example, a close-fitting elastic cap). Some electrode types, called active electrodes, have amplifiers built into the electrodes, making them more resistant to noise under certain conditions (but for a direct comparison between passive and active electrodes in different recording environments, see Laszlo, Ruiz-Blondet, Khalifian, Chu, & Jin, 2014). Conductive gel is used to establish the connection between the electrode and the skin. Especially with passive electrodes, light abrasion of the skin is typically used to establish a low impedance connection between the scalp and each electrode and to diminish skin potentials that can add noise to the recordings. The effectiveness of the electrode-to-scalp connection can be measured with an impedance meter; for passive electrodes, the impedance generally needs to be kept lower than for active electrodes (for the effects of impedance on data quality, see Kappenman & Luck, 2010).

The choice of how many electrodes to record from depends on the research question. Twenty to 32 electrodes are often sufficient for language processing studies that target broadly distributed components like the N400 and P600. Higher-density configurations, up to 256 channels, are also available, and the correspondingly improved resolution of the scalp topography can be advantageous for observing more focal effects and/or for modeling the underlying sources. There are trade-offs, however, as recording from more channels results in an increase in setup time, an increased probability of bridging between electrodes (resulting in a loss of separable signal), and a higher likelihood of a channel showing an artifact at any given time, leading to more data loss when rejecting entire trials from the analysis.

Another important choice is that of the reference electrode. Voltages express a *difference* in electrical potential between two points, meaning that at least two electrodes are necessary to measure a potential. EEG systems typically use a double subtraction ("differential amplification") involving a ground electrode and a reference electrode to reduce noise that the recordings from all electrodes have in common. Ideally, the reference electrode would be placed on an electrically neutral location on the body, but, in practice, no location is fully neutral. Most language studies make recordings relative to a reference electrode on the left mastoid, where there is a thick bone structure between the electrode and the brain, and re-reference to the average of left and right mastoid electrodes during analysis. However, some studies place the reference electrode on the nose or the earlobes. Still other studies convert the data to be able to use the average of all electrode sites as the reference ("average reference"). Because the choice of reference critically affects the amplitude and scalp distribution of the measured electrical signals, it is necessary to pay careful attention to the reference location when comparing datasets and best to follow the convention from other experiments within a subfield when designing a new study.

Electrodes are also placed on the face to help distinguish between artifacts and brain activity. Several types of artifacts stem from the eyes, because they act as an electrical dipole, causing large voltage changes during saccadic eye movements and blinks. Typically, electrodes are placed on the outer canthus of each eye to measure horizontal eye movements, which appear as square wave patterns of opposite polarity on each side of the eye. Another electrode is often placed on the infraorbital ridge below at least one of the eyes to measure blinks, which appear as large peaks of opposite polarity above and below the eye. For studies in which participants speak, it can be useful to place electrodes on the orbicularis oris muscle near the mouth, to monitor muscle activity, which is visible as bursts of higher frequency activity.

In the typical EEG setup, one computer presents the stimuli to the participant and another samples (digitizes) and stores the EEG data. The stimulus computer also sends brief event codes (also called triggers or markers) when stimuli are presented or responses are made to the digitization computer, to enable the later extraction of event-related data from the continuous EEG. The recorded signals are small, and need to be amplified considerably. They also need to be filtered, using an analog filter prior to digital sampling, in order to avoid aliasing, the phenomenon wherein activity at frequencies higher than half the sampling frequency (the Nyquist frequency) becomes misrepresented as lower frequency activity because it is sampled at a rate that is too low to reconstruct the original information. In practice, the EEG is low pass filtered at a frequency well below the Nyquist frequency.

EEG data always contain noise. There are external sources of noise, such as line noise from electrical devices near the participant. Most interference from electrical noise can be prevented by shielding the noise sources themselves (e.g., the monitor, cables) and/or by shielding the participant or recording devices (e.g., by seating the participant in a Faraday chamber). Nonetheless, many EEG recordings do contain some 60 Hz or 50 Hz line noise, depending on the country where the recordings are done. There are also physiological sources of noise, such as skin potentials, blinks, eye movements, and muscle activity. These are minimized by asking the participant to sit still and relax while fixating on the center of the screen. Many experimenters also ask participants to restrict blinking to certain points in the experiment, such as after every trial. Recordings are monitored by the experimenter in real time, so that they can detect excessive artifacts and other possible problems with the data, which is preferable to having to deal with them at the analysis stage.

Nature of Stimuli and Data

Many types of stimuli have been presented to participants while their EEG was recorded: written and spoken words, sentences, pictures, scenes, environmental sounds, and even video clips (e.g., Sitnikova, Kuperberg, & Holcomb, 2003). This range of stimuli permits language researchers to address all kinds of questions in comprehension, production, and acquisition. The EEG technique does put some constraints on the stimuli, however. In order to avoid eye movement artifacts, most studies present a single visual stimulus at a time, scaled to occupy a restricted part of the visual field. For instance, written sentences are usually presented word-by-word (although some groups have developed methods to record “fixation related potentials” during natural reading; e.g., Baccino & Manunta, 2005), and the presentation of auditory stimuli is usually combined with a constant visual stimulus, such as a fixation cross, to help keep participants' eyes on the center of the screen. Furthermore, each condition needs to have a relatively large number of stimuli; 30-60 for studies that target large components like the N400 and P600, more for studies that target smaller components (for discussion, see Luck, 2005). If it is difficult to design enough stimuli, or if the focus is on single-item ERPs, it is possible to compensate by testing more participants (Laszlo & Federmeier, 2011). The stimuli also need to be well controlled at many levels of analysis, because ERPs reveal the entire processing stream from perceiving, retrieving, evaluating, and (sometimes) responding to aspects of the stimulus. Full counterbalancing is optimal, but if this is not possible, stimuli can be matched on the relevant dimensions.

The task affects how the stimuli are processed. In some comprehension studies, participants make lexical decision responses, detect words, or answer comprehension questions. However, a strength of EEG as a continuous measure is that a task is not necessary in order to generate data. Thus, rather than requiring responses based on metalinguistic criteria, participants may simply be asked to read or listen for comprehension. This makes EEG experiments likely to capture the processes that listeners and readers also use outside the lab. Furthermore, it means that EEG can be used in populations for which behavioral testing is difficult, such as infants and certain patient groups. A particularly good example of this are Mismatch Negativity (MMN) studies asking when during development infants' speech perception system becomes more attuned to their native language versus other languages (Cheour et al., 1998). In production studies, classical picture naming tasks lend themselves to EEG investigations too (for review, see Ganushchak, Christoffels, & Schiller, 2011). However, the muscle artifacts generated by speaking are large and span a wide range of frequencies (Goncharova et al., 2003). This makes careful interpretation important, especially for later components close to articulation.

In all designs, a core aspect of the acquired EEG data is that they are multidimensional. They can be conceived of as a time sample × channel × trial matrix, with positive- and negative-going voltages. It is important to note that whether a signal is positive-going or negative-going in absolute terms does not allow for clear inferences about the underlying neurophysiological processes. The signal's polarity depends on the scalp electrode location: The same underlying brain activity that can be summarized as a current dipole will be measured as a positivity from one side and as a negativity from the opposite side. Moreover, ERPs are *relative* measures, recorded relative to a ground and reference channel and computed relative to a pre-stimulus baseline. Although the complexity of the data creates challenges for analysis, it is a key part of the utility of the technique, as it allows inferences not only about whether or not an experimental manipulation has an impact, but more specifically when and how. Such inferences can be especially strong when they involve well-characterized components linked to specific cognitive and neural functions. We have already seen an example of this way of exploiting the multidimensionality of the data: Whereas semantic and syntactic anomalies might both elicit longer response times relative to a congruent condition in a behavioral task, the distinct ERP effects (N400 and P600) these conditions elicit show that qualitatively different processes are recruited.

Collecting and Analyzing Data

A typical analysis pipeline for ERPs involves filtering, segmenting the epochs from the continuous data, baseline correction, artifact rejection, averaging, and statistical evaluation. Filtering, or reducing the presence of certain frequencies in the signal, is a large, complex topic, and beyond the scope of this chapter to adequately address. However, it is crucial that ERP researchers familiarize themselves with at least the basics (see Handy, 2004, and Luck, 2005, for useful discussion). There are high-pass filters (which let high frequencies pass through while attenuating lower frequencies), low-pass filters (which let lower frequencies pass through while attenuating higher frequencies), and band-pass filters (which combine low-pass and high-pass filters to let a frequency band pass through). Further properties of filters are the filter type (e.g., infinite impulse response, finite impulse response, each with various subclasses, such as Butterworth or Gaussian), the slope of the roll-off (which describes the steepness of the filter), and the frequency (defined as the half-amplitude cutoff or half-power cutoff). For ERPs, filtering is beneficial because it can reduce the amplitude of certain artifacts, facilitating the identification of ERP components and effects. High-pass filters can be used to reduce the influence of slow drifts and skin potentials in studies that do not target very slow components (which partly occupy the same frequency range). Low-pass filters can reduce the influence of high-frequency muscle activity. However, any filtering also leads to a loss of information and can distort the signal, which compromises the temporal resolution. High-pass filters, especially, can produce edge artifacts at the beginning and end of the piece of signal they are applied to. For this reason, high-pass filters are best applied to the continuous EEG, prior to segmentation.

To create ERPs, epochs around the onset of stimuli (or responses) of interest are extracted from the continuous EEG. A baseline correction is applied to each trial by subtracting the average voltage in the period preceding stimulus onset from all data points in the epoch, effectively setting the signal to zero at stimulus onset. This makes it easier to see the event-related modulations in the signal. Baseline periods in sentence comprehension paradigms are usually 100 to 200 ms long; short baselines minimize overlap with preceding events, whereas long baselines increase reliability of the estimate of baseline activity. There are a few studies that have filtered the signal instead of applying a baseline correction, as high-pass filtering can have similar effects as baseline correction when the cutoff frequency is fairly high and/or the filter is steep. However, as already mentioned, steep filters distort the signal (for discussion, see Luck, 2005). Under certain circumstances and settings, such filters can even make a P600 effect look like an N400 effect (Tanner, Morgan-Short, & Luck, 2015).

An important part of preprocessing is the removal of artifacts, including blinks, eye movements, muscle activity, drifts, and amplifier blocking (flatlining due to clipping, because the signal reached the end of the dynamic range of the amplifier). The identification of blinks is facilitated by subtracting the signal from electrodes above and below the eye (a vertical derivation), and the identification of saccades is similarly facilitated by computing a horizontal derivation of the signals from electrodes to the left and right of the eyes. Most studies reject trials that contain artifacts. Rejection decisions can be made using visual inspection, preferably while being blind to condition (although bias is unlikely, because the components of interest are usually not visible on individual trials). More common is a semi-automatic procedure in which one chooses participant-calibrated thresholds for automatic artifact detection methods (such as the maximal amplitude, peak-to-peak amplitude, or correlation with a step function). Instead of artifact rejection, which reduces the number of trials, artifact correction methods are also available. These methods measure or model the artifacts and remove them, for instance using independent components analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996). The non-artifactual independent components that ICA detects can also be studied as brain dynamics associated with cognitive processing, although researchers doing so will need a thorough understanding of the technique’s limitations, and it will be more difficult to compare the results of such statistically derived components with prior studies. In the next step, the artifact-free trials are averaged together point-by-point for each condition and each participant (or, in some studies, for each item). Finally, a grand average across participants is created to allow for visualization.

The participant averages are submitted to statistical analysis. Much ERP work relies on relatively straightforward statistical methods, validated by replication. Often, the research question is of the type “Does component X differ in amplitude between conditions?”, where the timing and scalp distribution of the component are known. This makes it possible to average across time points during which the component typically occurs and across the electrodes at which the effect tends to be maximal. The resulting values can be subjected to traditional analyses such as ANOVAs. To characterize ERP effects in terms of their scalp distribution, the locations of the electrodes or groups of electrodes on the scalp can be included as factors. Although the spatial resolution of ERPs is relatively poor compared with other neuroimaging techniques, a reliable difference between scalp distributions indicates that an experimental manipulation affected brain functioning, either by recruiting partially non-overlapping neuronal generators or by changing the amplitude of a shared generator. If the question is instead of the type “Does the *timing* of component X differ between conditions?” one can compute a fractional peak latency or fractional area latency measure. The fractional area measure computes the area under the curve within a time window and finds the point in time that divides the area into a specific fraction, such as 50% (Hansen & Hillyard, 1980). The fractional peak latency is calculated from the peak, back in time, as the point at which the signal reaches a particular fraction of the peak. Because noise makes the identification of peaks in individual subjects difficult, both of these measures benefit from applying them to “leave-one-out” grand averages, using the jackknife procedure (Miller, Patterson, & Ulrich, 1998; for recommended settings, see Kiesel et al., 2008) or, at minimum, from measuring peaks in low-pass filtered data.

In other types of experimental designs, the nature, timing, and distribution of the effects of interest are not known beforehand, and instead, the research question is of the type "Does the brain appreciate the difference between these conditions (and if so, how quickly)?". To handle such cases, data-driven "mass univariate" analyses have been developed and are implemented in, or compatible with, freely available software (e.g., Delorme & Makeig, 2004; Groppe, Urbach, & Kutas, 2011; Lopez-Calderon & Luck, 2014; Maris & Oostenveld, 2007; Oostenveld, Fries, Maris, & Schoffelen, 2011). Various approaches exist, but each share the advantage that the researcher need not specify a time window and set of electrode sites a priori. The first step of mass univariate approaches is to quantify the difference of interest in the form of some statistic (such as a *t* value) at each time point and each electrode. In a second step, a correction for multiple comparisons is applied, often based on permutation methods (or on the false discovery rate; Benjamini & Hochberg, 1995). Permutation procedures involve randomly swapping around the condition labels and re-running the statistical tests, and this process is repeated many times. Each permutation result contributes to a null distribution of test statistics, which acts as a benchmark for quantifying the size of effects that can occur simply by chance. Finally, the statistics from the actual (non-permuted) results are compared with the null distribution. If they are relatively "special" among the random permutations (i.e., if they are in a tail of the distribution), the difference between conditions is considered statistically significant. The main downside of these approaches is that they are less powerful compared with running an ANOVA or *t*-test directly on a time window. Thus, to avoid missing true effects, any a priori information that is available should be used to restrict the analysis and increase power. For instance, if one knows the distribution of an expected effect but not its timing, one can pick the electrode sites of interest but still test the entire epoch point-by-point – or vice versa. The test results allow the researcher to inspect at which time points and electrodes any differences between the conditions occurred, although the extent to which this time course can be interpreted as onsets and offsets of effects depends on the multiple comparisons procedure. For instance, the cluster-based permutation approach only tests the general null hypothesis that there is no difference between the conditions (the conditions are exchangeable); the false alarm rate is not controlled at the level of the onsets and offsets of clusters (Maris, 2012). Taken together, there are suitable statistical methods for most designs and extents of a priori knowledge.

An Exemplary Study

To help make the above more concrete, we discuss an example study by Van Petten, Coulson, Rubin, Plante, and Parks (1999), who used ERPs to investigate spoken-word comprehension in sentence context. The study used speech, which is less often used in ERP studies than written words (as visual stimuli are easier to time-lock to), and capitalized on several advantageous features of the ERP method in its experimental design.

Spoken language input unfolds over time and lacks clear cues to word boundaries, unlike alphabetic written text in which the spaces help. Listeners activate multiple candidate words (like “cat”, “a” and “log” while hearing “catalog”) in an incremental fashion based on incomplete input (Marslen-Wilson & Welsh, 1978). Van Petten et al. investigated the extent to which the meanings of these candidate words are activated and when and how they make contact with sentence context.

Out of context, a word can be identified as soon as the acoustic input becomes uniquely consistent with that word. This point in time is known as the isolation point and it can be empirically established using the gating task (Grosjean, 1980), in which listeners are presented with successively longer onset fragments of the word and asked to guess what the word is or is going to be. As the fragments get longer, listeners converge on the same response. In supportive sentence contexts, the responses converge earlier, with less acoustic input (Grosjean, 1980). Some studies used cross-modal priming paradigms, in which participants make lexical decisions to visually presented words while listening to words in context, to investigate the semantic activation of word candidates (Chwilla, 1996; Moss & Marslen-Wilson, 1993; Zwitserlood, 1989). For example, while hearing successively longer fragments of the word "generous" in a supportive sentence context that is inconsistent with "general", participants would be probed with "gift" (associated with the contextually supported word) and "army" (assessing activation of the contextually unsupported, but initially overlapping "general"; Zwitserlood, 1989). However, the results were mixed, and the nature and time course of the processes between hearing the fragment, seeing the target, and making a response was not known.

Van Petten et al. (1999) used the N400 effect to examine the initiation of semantic processing relative to the isolation point. The study focused directly on processing of the spoken word itself. If context-dependent semantic processing of words only begins after they have been fully recognized, the N400 to words that fit and words that do not fit in the sentence should only begin to differ after the isolation point. However, if semantic processing begins to operate on incomplete input, then the N400 effect could begin prior to the onset of the isolation point, as soon as the acoustic input diverges from any contextual expectations that listeners might have. Participants listened to sentence contexts like "It was a pleasant surprise to find that the car repair bill was only seventeen…", which ended in a word that fit in the context ("dollars"; *cohort congruous* condition), an incongruous word that rhymed with the congruous word ("scholars";

rhyme condition), or in an incongruous word that shared initial phonemes with the congruous word ("dolphins"; *cohort incongruous* condition).

Figure 2 shows the results (of Experiment 3, continuous speech; not shown are Experiment 1, a gating study, and Experiment 2, which presented a pause before the final word).

< Insert Figure 2 about here >

In the ERPs time-locked to word onset, both incongruous conditions elicited much larger N400 amplitudes than the congruous condition. This replicated previous studies showing how contextual support reduces N400 amplitude. Comparing the incongruous conditions, however, there was a large difference in onset timing of the N400. The semantically incongruous words that shared initial phonemes with the congruous completion (Cohort incongruous) elicited an N400 that was delayed by about 200 ms compared with those that did not share initial phonemes (Rhyme). These results already suggest that the isolation point may not be a crucial determinant of N400 onset, but to correct for variability in isolation point across individual words, the ERPs were also time-locked to the isolation point. When the incongruous words shared initial phonemes with the congruous word, the N400 onset occurred at the isolation point. But when the incongruous word had different initial phonemes, the N400 onset occurred ~200 ms prior to the isolation point. This strongly demonstrates that context-driven semantic processes do not wait until the acoustic signal has fully disambiguated the word. Instead, the results argue for a continuous mapping from acoustic input to semantic representations. Note that the semantic interpretation of these results is afforded by the ability to identify the pre-isolation-point ERP effect as being on the N400 rather than some other component (for discussion about a phonological mismatch component, see Connolly & Phillips, 1994; van den Brink, Brown, & Hagoort, 2001). Van Petten et al. made this argument based on the waveform characteristics and functional sensitivity of the effect, pointing out as well that there was no evidence for additional components – no additional peaks in individual subject ERPs and no shift in scalp distribution over time.

The advantages of ERPs for addressing the questions of interest in this study are clear. The experimental design made use of the fact that the EEG signal is an instantaneous and continuous reflection of how the speech signal is processed, obviating the need to make inferences based on downstream consequences and metalinguistic judgments. The study also exemplifies the utility of time-locking to different parts of the speech signal, in this case allowing for the investigation of context effects separately at points in time before and after any purely context-independent word recognition processes could have disambiguated the input.

Problems and Pitfalls/Advantages and Disadvantages

This section discusses challenges with ERP methods, as well as how some of these issues are being overcome. One fundamental challenge that has already been discussed is that the EEG contains high levels of noise, necessitating techniques for extracting a stable signal of interest – most commonly, averaging. However, as with any average, an average ERP may not accurately reflect the processing pattern in individual participants or on individual trials. For instance, a decrease in amplitude in one condition relative to another could be due to a component being attenuated on every trial, on only a subset of trials, or even as a result of latency variation, such that the timing of the component is more variable in one condition than in the other, leading to a reduced amplitude in the average (Spencer, 2004). Furthermore, the ERP from a given study may contain a biphasic N400-P600 pattern of effects when averaged, but this could in principle stem from a combination of some trials (and/or, in the grand average, participants) with only an N400 modulation and some with only a P600 modulation. Moreover, ERP datasets are not unlikely to be somewhat unbalanced in terms of the number of trials and the identity of the items going into each condition average, because of artifact rejection and, in some designs, binning that is based on participants' behavioral response patterns. Although this is unlikely to affect outcomes in experiments wherein the same perceptual stimuli are rotated across conditions and only a random 5-10% of the trials are rejected, sometimes the question addressed necessarily contrasts different items, such as in word recognition experiments that try to discern the effects of various psycholinguistic variables.

To address such concerns, ERP researchers have begun to use alternative statistical methods that have also gained popularity in the behavioral and eye-tracking literatures, such as mixed-effects regression models (e.g., Baayen, Davidson, & Bates, 2008). Instead of averaging, mixed-effects (or hierarchical) models directly model the trial-level data. This allows for the simultaneous inclusion of participants and items as random factors, which makes it possible to include any measured participant characteristics (such as working memory capacity) and item characteristics (such as word frequency) and to examine effects of practice or fatigue across trials. In principle, estimating ERPs by running a regression model at the level of individual trials is not dissimilar to averaging. However, whereas averages can be distorted in unpredictable ways by unbalanced missing data, mixed-effects models can deal with missing data in a principled way because at the individual trial level, it is known by which participant and item a (brain) response was elicited. Although the field has not yet settled on conventions regarding the various possible ways of modeling multiple electrodes and time points/windows, there are promising applications of trial-level analyses to ERPs, including the investigation of how continuous predictors such as word position in a sentence affect ERPs (Payne, Lee, & Federmeier, 2015), of non-linear relationships between predictors and ERPs (Tremblay & Newman, 2015), and of ways to handle overlapping responses to distinct events (Smith & Kutas, 2015).

Another disadvantage of averaging is that it does not capture certain aspects of the EEG signal. For activity to show up in an average ERP, it needs to be not only time-locked to an event, but also phase-locked to it; that is, the peaks and troughs in the waveform need to be aligned in time across different trials. Such "evoked" activity can be contrasted with "induced" activity, which is time-locked but not phase-locked. Even though its amplitude can be large, non-phase-locked activity is unlikely to become visible in the ERP because the peaks have a variable latency relative to the stimulus and largely cancel each other. The success of ERPs in delineating core cognitive processes suggests that phase-locked activity captures something fundamental about cognition and brain functioning. However, current views of brain functioning also emphasize the role of oscillatory activity (which is often not phase-locked) in critical aspects of cognitive processing, including language (for discussion about oscillatory activity as the EEG signature of the coupling and uncoupling of neuronal networks, see Bastiaansen, Mazaheri, & Jensen, 2008; see also Buzsáki, 2006).

Therefore, a growing number of language processing studies employ time-frequency analysis to make visible not only phase-locked but also non-phase-locked activity, as in other fields in which these analyses are routinely applied. Time-frequency analysis involves decomposing the EEG signal into multiple frequencies and quantifying power (amplitude squared) at each frequency over time. The analysis is applied to individual trials and then an average across trials is taken. Different frequency bands that respond differently to cognitive manipulations have been identified and labeled: delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (>30 Hz). The frequency bands are not fixed but merely serve as a guideline to facilitate communication. Peak alpha frequencies, for instance, actually differ between participants, as well as between tasks within the same participants (Haegens, Cousijn, Wallis, Harrison, & Nobre, 2014; Klimesch, 1999).

< Insert Figure 3 about here >

Various time-frequency analysis methods are commonly used, including the short-time Fast Fourier Transform (FFT), Morlet wavelet analysis, and filtering combined with the Hilbert transform (for discussion, see Cohen, 2014). Each of these methods has its own parameters, but when the parameter settings are matched, the results look similar; in fact, the three approaches are mathematically equivalent to one another (Bruns, 2004). As shown in Figure 3, the result from such an analysis can be visualized as a spectrogram, with time on the x-axis, frequency on the y-axis, and color coding for increases and decreases in power at the different frequencies over time. It is important to note that these spectrograms do not have the temporal resolution that ERPs have; there is considerable temporal and frequency "smearing". In signal processing, there is an inverse relationship between frequency precision and temporal precision, and this trade-off is determined by the analysis parameters (such as the number of wavelet cycles, the filter settings when using the filter-Hilbert method, or the FFT window length and taper properties). For instance, when using a 400 ms moving-window FFT approach, each "pixel" in the spectrogram is calculated using the data from 200 ms before and 200 ms after the pixel (although data points closer to -200 and +200 ms will have progressively less influence, depending on the shape of the taper used). Using a larger time window would improve the frequency precision at the expense of temporal precision, whereas using a smaller window would improve the temporal precision at the expense of frequency precision. In general, by decomposing the signal into its constituent frequencies, some temporal resolution is always sacrificed.

Compared with the rich and well-established literature on ERPs, less is currently known about the role of non-phase-locked activity in language processing. This will likely change in the coming years, but it has important implications for statistical analysis. With ERPs, one can target a particular component with a known latency and scalp distribution and reduce the data accordingly for analysis. With time-frequency approaches, it is more often the case that the latency, scalp distribution, and frequency bands in which effects will occur are not known before inspecting the data. Thus, in this case it becomes especially important to consider using data-driven statistical methods that deal with the problem of multiple comparisons, such as the ones discussed in the section Collecting and Analyzing Data, which can straightforwardly incorporate frequency as an additional dimension besides time and space (Maris & Oostenveld, 2007).

Another challenge with ERPs and EEG in general is that it is difficult to infer which brain areas were active based on just the scalp topography of a component or effect. For most psycholinguistic questions, the timing of the brain activity is probably more germane than its source location. But when localizing activity is important, and one does not want to sacrifice temporal resolution (as occurs with functional magnetic resonance imaging (fMRI)), one can turn to magnetoencephalography (MEG).

MEG is similar to EEG in many ways (for a detailed introduction to the method, see Hämäläinen et al., 1993). The same types of neural processes that produce the electrical activity reflected in the EEG also produce the magnetic activity visible in the MEG. Both power changes and event-related fields (ERFs, the magnetic equivalent of ERPs) can be analyzed. Many ERP components have a magnetic counterpart. In those cases, the corresponding MEG components are generally named like the ERP ones, with an “m” appended to the label. For example, the N400m is the MEG response taken to reflect activity shared with the N400 (e.g., Halgren et al., 2002; Simos, Basile, & Papanicolaou, 1997). As with EEG, the temporal resolution of MEG is a major strength.

Despite these similarities, there are important differences between EEG and MEG. The MEG signal is recorded using superconducting quantum interference devices (SQUIDs), which are highly sensitive magnetometers that need to be cooled in liquid helium at a very low temperature (4 Kelvin). Gradiometers, which measure the difference between two or more neighboring coils, make the signal especially sensitive to nearby brain sources and decrease the influence of more distant noise sources, including the heart. Most current MEG systems contain several hundred gradiometers, arranged in a helmet-like shape. Because the brain signals are much weaker than magnetic noise coming from, for example, radios, moving cars and elevators, MEG systems are usually placed in a magnetically shielded room. Both the initial purchase of the MEG system and the necessary supplies of liquid helium make the method considerably more costly than EEG.

One of the main virtues of MEG stems from the fact that, compared with electrical signals, magnetic signals are less spatially smeared by the skull between the brain and the sensors (e.g., Hämäläinen et al., 1993). Skin potentials, which complicate EEG recordings at low frequencies, are also not seen by MEG. Furthermore, certain widespread muscle artifacts in the EEG may be reduced in the MEG, which can facilitate the study of speech production (Hari & Salmelin, 2012; for examples, see Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998;

Salmelin, Hari, Lounasmaa, & Sams, 1994). Thus, certain types of distortion and noise are more problematic for EEG than MEG. At the same time, MEG is sensitive to a different subset of brain signals: Whereas currents that are oriented tangentially to the skull (in the walls of cortical sulci) are seen by both MEG and EEG, currents that are oriented radially to the skull (as on gyri, encompassing an estimated one third of the brain's cortical surface) are seen by EEG only. Magnetic signals, compared to electrical ones, also show a steeper decline with distance, making MEG relatively more selective to superficial brain sources. Source localization with MEG is thus easier because a more restricted subset of brain activity is being modeled.

Sources can be modeled using various methods, including an equivalent current dipole, multiple dipoles, or beamforming techniques (for discussion, see Hari & Salmelin, 2012). In each case, certain assumptions are necessary, because the "inverse problem" has no unique solution (multiple source configurations can generate the same scalp distribution). Incorporating an anatomical MRI scan into the analysis can further help reduce the source modeling solution space (this is true for source modeling with EEG as well). Some current MEG studies go beyond localization and use sophisticated connectivity methods at the source level to investigate communication between different brain areas (for review, see Bressler & Seth, 2011; David et al., 2006; Schoffelen & Gross, 2009).

Overall, a broad characterization of the two methods is that MEG usually sees less than the EEG sees, but sees it more clearly (Cohen & Halgren, 2009). However, it is perhaps most useful to view these methods as complementary, and, indeed, some have argued that the best source localization will come from combined EEG and MEG (Cohen & Halgren, 2009; Sharon, Hämäläinen, Tootell, Halgren, & Belliveau, 2007).

In summary, this chapter discussed how the noninvasive measurement of electrical brain activity generates some of the most direct evidence regarding the processes underlying language comprehension, production, and acquisition in the brain. The established approaches, supplemented by current developments, are likely to continue to provide important new insights that keep challenging our views of cognition and brain functioning.

Acknowledgement

This work was supported by a James S. McDonnell Foundation Scholar Award and NIH grant AG026308 to K. D. F.

Key Terms

EEG: electroencephalogram, the record of electrical brain potentials.

MEG: magnetoencephalogram, the record of magnetic brain potentials.

ERPs: event-related potentials, waveforms averaged across multiple trials time-locked to an event.

ERP component: one of the component waves of the ERP waveform.

ERP effect: an experimentally isolated difference between conditions, often a modulation of an ERP component.

References

Baayen, R.H., Davidson, D. J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baccino, T., & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, *19)*, 204-215.

Bastiaansen, M. C. M., Mazaheri, A., & Jensen, O. (2008). Beyond ERPs: Oscillatory neuronal dynamics. In S. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components*. New York: Oxford University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. *Series B (Methodological), 57*, 289-300.

Bressler, S. L., & Seth, A. K. (2011). Wiener–Granger causality: A well established methodology. *Neuroimage, 58*, 323-329.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127-143.

Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods, 137*, 321–332.

Buzsáki, G. (2006). *Rhythms of the brain.* New York: Oxford University Press.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*, 351–353.

Chwilla, D. J. (1996). *Electrophysiology of word processing: The lexical processing nature of the N400 priming effect.* Nijmegen University.

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice.* Cambridge, MA/London, UK: MIT Press.

Cohen, D., & Halgren, E. (2009). Magnetoencephalography. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience, volume 5*, 615-622.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. Journal of *Cognitive Neuroscience*, *6*, 256-266.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21-58.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., & Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, *30*, 1255-1272.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8,* 919-24.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9-21.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259-281.

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology*, 2, 208.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: Spectral and topographical characteristics. *Clinical Neurophysiology, 114,* 1580-1593.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*, 1711-1725.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*, 267-283.

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *Neuroimage, 92*, 46-55.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-483.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*, 1101-1116.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., Lounasmaa, O. (1993). Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65,* 1-93.

Handy, T. C. (Ed). (2004). *Event-related potentials: A methods handbook.* Cambridge, MA: MIT Press.

Hansen, J. C., & Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49,* 277-290.

Hari, R., & Salmelin, R. (2012). Magnetoencephalography: From SQUIDs to neuroscience: Neuroimage 20th anniversary special edition. *Neuroimage, 61*, 386-396.

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology, 47,* 888–904.

Kiesel, A., Miller, J. O., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology, 45,* 250-274.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29,* 169-195.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience,* 5, 196-214.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language, 100,* 257-261.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17,* 117-129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4,* 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48,* 176-186.

Laszlo, S., Ruiz-Blondet, M., Khalifian, N., Chu, F., & Jin, Z. (2014). A direct comparison of active and passive amplification electrodes in the same amplifier system. *Journal of Neuroscience Methods*, 235, 298-307.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10,* 553-567.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 1-14.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luck, S. J., & Kappenman, E. S. (Eds.). (2011). *The Oxford handbook of event-related potential components*. New York: Oxford University Press.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems, 8,* 145-151.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*, 549-565.

Maris E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177-190.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*, 29-63.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology, 35*, 99–115.

Millett, D. (2001). Hans Berger: From psychic energy to the EEG. *Perspectives in Biology and Medicine*, 44, 522-542.

Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1254-1276.

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217-226.

Näätänen, R., Lehtoskoskl, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Nunez, P. L., & Srinivasan, R. (2006). *The electric fields of the brain: The neurophysics of EEG.* Oxford: Oxford University Press.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.

Osterhout, L., & Holcomb, P.J . (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*, 1456-1469.

Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, *368*, 463-465.

Schoffelen, J. M., & Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, *30*, 1857-1865.

Sharon, D., Hämäläinen, M. S., Tootell, R. B., Halgren, E., & Belliveau, J. W. (2007). The advantage of combining MEG and EEG: Comparison to fMRI in focally stimulated visual cortex. *NeuroImage, 36,* 1225–1235.

Simos, P. G., Basile, L. F., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain research*, *762*, 29-39.

Sitnikova, T., Kuperberg, G.R., & Holcomb P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40,* 160-164.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157-168.

Spencer, K. M. (2004). Averaging, detection and classification of single-trial ERPs. In Todd C. Handy (Ed.), *Event-related potentials. A method handbook*. Cambridge, MA: MIT Press.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement. *Cortex, 66,* 149-155.

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*, 997-1009.

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology, 52*, 124-139.

Van Den Brink, D., Brown, C., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967-985.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25,* 394-417.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 787-806.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25-64.

Figure captions

Figure 1. Idealized example of an event-related potential waveform in response to a visual stimulus, with labeled positive and negative peaks. A single channel is shown; negative is plotted up.

Figure 2. Grand average ERPs from three parietal channels, elicited by the final words in the three conditions. In the left column, time zero is the onset of the word. In the right column, time zero is the isolation point. Source: Van Petten et al., 1999, Fig. 10.

Figure 3. Simulated EEG data illustrating the difference between ERPs and time-frequency analyses in their sensitivity to phase-locked (evoked) and non-phase-locked (induced) activity. The first response is time-locked and phase-locked to time zero, whereas the second response is time-locked but not phase-locked. The first response shows up both after ERP averaging (as an oscillation) and after time-frequency analysis of power (as a power increase at around 10 Hz). The second response is canceled by ERP averaging, but is preserved in time-frequency analysis of power. Figure taken from Bastiaansen, Mazaheri, and Jensen (2008).

Figure 1

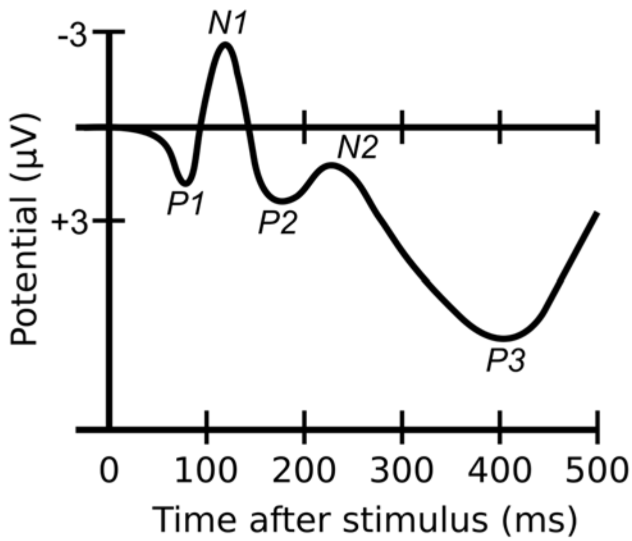

N1
-3
N2
P1
+3
P2
Potential (µV)
P3
0
100
200
300
400
500
Time after stimulus (ms)

Figure 2

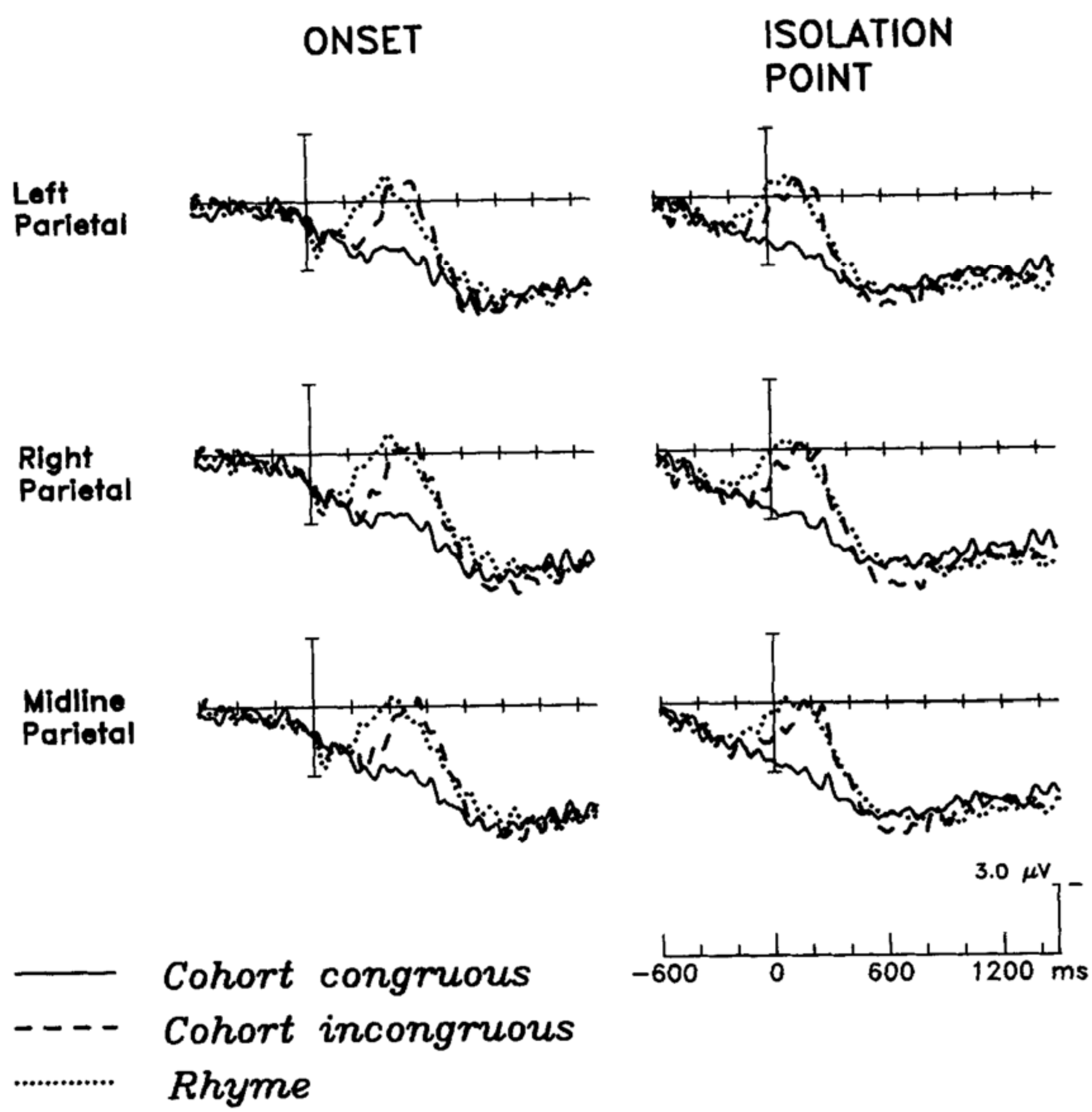
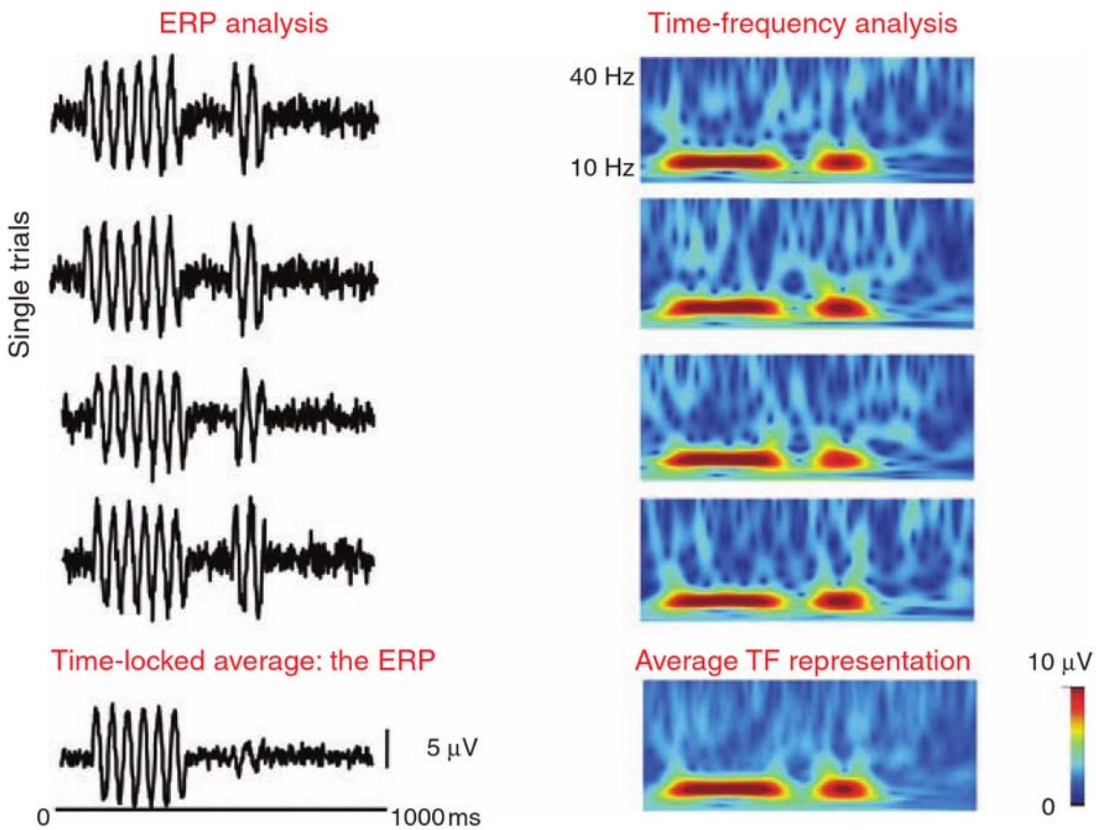

ONSET
ISOLATION POINT
Left Parietal
Right Parietal
Midline Parietal
3.0 μV
-600
0
600
1200
ms
Cohort congruous
Cohort incongruous
Rhyme

Figure 3



ERP analysis
Time-frequency analysis
40 Hz
10 Hz
Single trials
Time-locked average: the ERP
Average TF representation
10 μV
5 μV
0
1000 ms
0